

3

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
9 January 2003 (09.01.2003)

PCT

(10) International Publication Number  
**WO 03/002738 A1**

(51) International Patent Classification<sup>7</sup>: C12N 15/10, 15/76, 1/21

(21) International Application Number: PCT/GB02/02884

(22) International Filing Date: 24 June 2002 (24.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0115894.8 28 June 2001 (28.06.2001) GB

(71) Applicant (for all designated States except US): PLANT BIOSCIENCE LIMITED [GB/GB]; Norwich Research Park, Colney Lane, Norwich, Norfolk NR4 7UH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): FOWLER, Kay [GB/GB]; 20 Fifers Lane, Old Catton, Norwich, Norfolk NR6 7AF (GB). KIESER, Tobias, Edwin [GB/GB]; 7 The Hedgerows, Norwich, Norfolk NR5 9BP (GB).

(74) Agents: DENISON, Christopher, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London, Greater London WC2B 6HP (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 03/002738 A1

(54) Title: METHODS AND MATERIALS FOR GENERATING GENETIC DISRUPTIONS IN BACTERIAL CELLS

(57) Abstract: Methods and materials for generating genetic disruptions in bacterial host cells, especially *Streptomyces* spp., and more particularly for generating libraries of bacterial host cells having respective such disruptions. The methods involve conjugation into the host cell of a plasmid bearing a transposable element which comprises a marker.

Methods and Materials for Generating Genetic Disruptions  
in Bacterial Cells

The present invention relates to methods and materials for generating genetic disruptions in bacterial host cells, especially *Streptomyces* spp., and more particularly (but not necessarily) for generating libraries of bacterial host cells having respective such disruptions.

5      The most preferred bacterial host cells are cells of *S.coelicolor* M145, a plasmid-free ( $SCP1^-$   $SCP2^-$ ) derivative of the wild type *S.coelicolor* A3(2) strain. *Streptomyces coelicolor* A3(2) is genetically the most studied *Streptomyces* species. It is for this reason that the

10     entire single, linear, 8,800 kb chromosome of *S.coelicolor* A3(2) was sequenced (using *S.coelicolor* M145) at the Sanger Centre. 8000 genes were predicted, with an average gene size of 1.1 kb. 53% of these 8000 genes have no known function. Most of these genes are probably non-  
15     essential for growth under normal laboratory conditions. It is of great interest to study these genes, and to this end it is of interest to construct a library of mutants containing disruptions in different non-essential genes (and/or their control sequences), resulting in knock-out  
20     mutations of those genes.

25     A library of mutant host cells should ideally be representative (i.e. including knock-out mutants for as many non-essential genes as possible, preferably substantially all non-essential genes), with as little redundancy as possible (i.e. avoiding as much as possible the presence of multiple knock-out mutants of the same gene). The mutations should ideally be random (i.e. without bias towards particular target nucleic acid

sequences, which might lead to a non-representative and/or redundant library), independent, and associated with a marker ("tagged") to allow identification of mutants.

5 Transposon mutagenesis has been shown previously to offer certain advantageous features in the generation of untargeted mutations. The transposon should transpose *in vivo*, randomly, and efficiently, and insertions should be stable. Transposon insertions are expected to cause  
10 complete and unconditional gene inactivation, therefore essential genes will not be represented in a mutant library.

Chung (1987) introduced a viomycin resistance gene (*vph*) from *Streptomyces vinaceus* into Tn4556, a naturally occurring *Streptomyces fradiae* transposon (GenBank accession number M29297). Tn4556 is cryptic, in that it does not encode a marker (see Figure 1), which restricts its use as a genetic tool. The recombinant, tagged, 20 transposon was designated Tn4560 (Figure 1).

Tn4556 is a Tn3-like transposon, based on sequence identity (Olson and Chung, 1988; Siemieniak et al., 1990). Tn3-like transposable elements transpose, one per 25 replicon, by a replicative transposition mechanism (the donor molecule remains intact after transposition). Transposition generates a 5 bp direct repeat of the insertion site.

30 Tn4560 was delivered into *Streptomyces* spp. on the 14.3 kb temperature-sensitive plasmid pUC1169 (pMT660::Tn4560; Chung, 1987) (see Figure 2). pMT660 (Birch and Cullum, 1985) is a temperature-sensitive derivative of pIJ702 (Katz et al., 1983), itself a derivative of the naturally

occurring *S. lividans* plasmid pIJ101 (Kieser et al., 1982; GenBank accession number M21778). pUC1169, in addition to the *vph* marker carried by the transposon, also has a thiostrepton resistance marker (*tsr*), in the pMT660 portion of the plasmid.

Despite demonstrating effective transposition from plasmid pUC1169 of transposon Tn4560 into *Streptomyces* host cell chromosomes, Chung et al. have not reported its use to generate a library of mutant host cells. This may be due at least in part to certain laborious or inconvenient steps which must be performed for transposon mutagenesis using this plasmid, which hinder the transposon mutagenisation of large numbers of host cells.

In the work upon which the invention is based, the inventors investigated conjugal transfer of a plasmid containing Tn4560 from *Escherichia coli* to *S. coelicolor* M145 as a method for transposon delivery and construction of a Tn4560 mutant library. In order to achieve this, a conjugative *E. coli*-*Streptomyces* shuttle vector that contains Tn4560 and that does not replicate once transferred to *S. coelicolor* M145 was constructed. Such a vector avoids two laborious or inconvenient and potentially mutagenic procedures: protoplast transformation and regeneration and incubation at elevated temperatures.

Using this method, the inventors have shown that transposition of Tn4560 occurred randomly around the chromosome of *S. coelicolor* M145, and that 90% of insertions fell within coding sequences. Analysis of the target site sequences of Tn4560 insertions did not reveal

any obvious nucleotide sequence similarities or preference for DNA of a particular %G+C content.

Tn4560 transposes from plasmid to chromosome at a frequency between  $10^{-3}$  and  $10^{-4}$  colony forming units (cfu) (Chung and Crose, 1990), although higher apparent transposition frequencies have been reported for transposition into SCP2<sup>\*</sup> plasmid derivatives (Brolle et al., 1993).

10

Transposon mutagenesis has therefore been demonstrated to be a successful and efficient method for generating a mutant library in *S.coelicolor* M145.

15 Accordingly, in a first aspect, the present invention provides a plasmid for generating a genetic disruption in a bacterial host cell, the plasmid comprising:  
a transposable element encoding functions to enable transposition of the transposable element into the host  
20 cell nucleic acid, the transposable element further comprising a first marker gene; and  
an origin of transfer.

The provision of the origin of transfer in the nucleic acid construct allows the introduction of the plasmid into the host cell by conjugation from a donor cell.

Preferably the origin of transfer is an *oriT* which can be mobilised by the helper plasmids pUZ8002 and pUB307, such 30 as an *oriT* from an IncP-group plasmid, such as RP4 (also designated RP1/RK2; Pansegrouw et al., 1994), preferably having the nucleic acid sequence:

CCGGGCAGGA TAGGTGAAGT AGGCCACCC GCGAGCGGGT  
GTTCCCTTCTT CACTGTCCCT TATTGCACC TGGCGGTGCT  
CAACGGGAAT CCTGCTCTGC GAGGCTGGC,

or a variant thereof having origin of transfer function.

5 However, the use of any other suitable origin of transfer  
is also contemplated.

Preferably the plasmid comprises an origin of replication  
which is functional in the donor cell. The skilled person  
10 will be able to choose an appropriate origin of  
replication using standard reference texts, as well as by  
referring to the Examples.

The host cell is preferably an actinomycete cell, more  
15 preferably a streptomycete cell, more preferably a cell of  
the genus *Streptomyces*, more preferably a cell of the  
species *S.coelicolor*, more preferably a cell of the strain  
*S.coelicolor A3(2)*. It is preferably a cell which lacks  
plasmids that accept transposons such as *Tn4560*, e.g. for  
20 *S.coelicolor A3(2)* a cell which lacks the naturally  
occurring *SCP1* and/or *SCP2* plasmids, e.g. strain *M145*,  
which lacks both *SCP1* and *SCP2*.

For certain embodiments, which use an actinomycete  
25 bacterial host cell, the present invention provides the  
advantages that introduction of the transposon-bearing  
plasmid can be accomplished without protoplast  
transformation and regeneration, using conjugation into  
pre-germinated spores, and that (especially with  
30 conjugation from *E.coli*) the procedure is broadly  
applicable in introducing DNA into actinomycetes other  
than *S.coelicolor* (Matsushima et al., 1994). As well as  
being laborious, protoplast transformation and  
regeneration procedures may produce mutations.

The donor cell is preferably of a different cell type from the host cell; any convenient bacterial cell may be used. For convenience, however, the donor cell is most 5 preferably an *E.coli* cell.

Where host cell has methylation-specific restriction system (e.g. *S.coelicolor* has such a system, although the related strain *S.lividans* does not, MacNeil et al., 1992), 10 the donor cell is preferably methylation-deficient, e.g. *E.coli* strain ET12567 (MacNeil et al., 1992).

The transposable element is preferably one which undergoes replicative transposition (e.g. a Tn3-like transposable 15 element), thereby leaving the plasmid intact after transposition. However, it is contemplated that transposable elements which undergo transposition by simple insertion (e.g. a Tn10-like transposable element) may also be used.

20 The skilled person is well aware of the structure of transposable elements, which have terminal inverted repeats (though these need not be perfect repeats) and a transposase gene, the product of which is thought to 25 recognise the terminal inverted repeats.

Preferably the transposable element also comprises a resolvase gene and internal resolution site, which combine with the transposase for replicative transposition. It 30 is, however, thought that resolvase function may alternatively be provided by the host cell (this function can be provided by the *E.coli recA* system).

The transposable element is preferably one which confers transposition immunity following transposition into a replicon (e.g. a Tn3-like transposable element). This prevents the plasmid from causing more than one 5 transposition event per replicon. In preferred embodiments, in which the host cell (e.g. *S.coelicolor* M145) lacks plasmids which accept the transposon, this means that there can only be one transposition event per cell.

10

The transposable element is preferably one which does not demonstrate high target site specificity. This provides the advantage of allowing essentially untargeted gene disruption to occur, so that in a library of mutant host 15 cells, different disruptions will be expected to be present in different host cells, rather than bias towards disruption at a limited number of locations possessing the relevant specific target site. The Tn3 and IS1 transposable elements show preference towards AT-rich 20 regions. Such target site specificity is not regarded to be high for the present purposes and transposable elements having Tn3-like or IS1-like inverted repeat sequences are preferred for this reason. However, there are hotspots 25 for Tn3 transposon insertion, i.e. locations in host cell nucleic acid which have homology with the Tn3 inverted repeat sequences, at which insertion occurs with greater frequency than at other locations, so some bias towards these hotspots might be expected. Transposons having such inverted repeat sequences are not especially preferred, to 30 avoid bias in the site of genetic disruptions in host cell libraries. Especially preferred transposons therefore have inverted repeat sequences which do not display a significant preference for particular insertion hotspots, e.g. Tn10-like inverted repeat sequences.

Tn4556-like or Tn4556-derived transposable elements are particularly preferred for this reason. The transposable element Tn4560 disclosed herein, and transposable elements  
5 derived therefrom, are especially preferred.

The transposase gene of the transposable element is preferably from the same source as the terminal inverted repeat sequences.

10

While the skilled person will be aware of many different transposable elements and will be able to choose or produce suitable examples for use in the present invention (see e.g. Singleton and Sainsbury (1987) Dictionary of

15 Microbiology and Molecular Biology, 2nd edition, John Wiley & Sons, under the entries "transposable element", "transposition immunity" and other entries referred to therein, as well as references cited in those entries, e.g. the review Grindley and Read (1985); see also Berg  
20 and Howe (1989) and Kieser et al (2000)), the transposable element is preferably Tn3-like, based on sequence and structural similarity to the Tn3 transposon. More preferably the transposable element is derived from a known such Tn3-like transposable element. See the entry  
25 under Tn3 in Singleton and Sainsbury, and Shapiro (ed) (1983) "Mobile Genetic Elements", Academic Press.

For the avoidance of doubt, it is hereby stated that the transposable element is preferably a DNA-only transposon,  
30 rather than a retrotransposon (e.g. a retroviral-like retrotransposon or a nonretroviral retrotransposon). For a description of these different classes of transposable element, see e.g. Alberts et al. (2002) Molecular Biology of the Cell, 4th edition, Garland Science, New York.

The transposon Tn4556 was identified as being Tn3-like by virtue of displaying the following sequence identity with the *Tn3* transposon: for the transposase gene, 66% identity at the deduced amino acid level (over 892 amino acids); for the resolvase gene, 45% identity at the deduced amino acid level (over 185 amino acids); and for the terminal inverted repeat sequences, 70% identity at the nucleic acid level (over 38 bp). Preferred transposons have similar levels of identity with *Tn3*, or more preferably with *Tn4556*. For example, the transposons of the invention preferably comprise a transposase gene having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity at the amino acid level with the transposase of *Tn3* or, more preferably, *Tn4556*. Similarly, the transposons of the invention preferably comprise a resolvase gene having at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity at the amino acid level with the resolvase of *Tn3* or, more preferably, of *Tn4556*. Similarly, the transposons of the invention preferably comprise terminal inverted repeat sequences having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity at the nucleic acid level with the terminal inverted repeat sequences of *Tn3* or, more preferably, of *Tn4556*. The sequence of *Tn3* is available under GenBank accession number V00613 (version V00613.1; GI:43710) and the sequence of *Tn4556* is available under GenBank accession number M29297 (version M29297.1; GI:153509). The resolvase gene of *Tn3* is identified as "repressor" in GenBank accession number V00613, as it serves the dual role of encoding the resolvase and repressing the resolvase and transposase.

genes (which repression is involved in transposition immunity).

Preferably the plasmid comprises, outside the transposable element, a second marker gene which differs from the first marker gene (which is within the transposable element). Loss of the second marker, but retention of the first marker, is indicative of a transposition event.

10 The marker genes preferably encode selectable markers and more preferably are antibiotic or heavy metal resistance genes. Where the transposon is derived from a naturally occurring transposable element which lacks a first marker gene, the first marker gene may be introduced

15 recombinantly.

Preferably the plasmid is non-replicable in the host cell. For example, the plasmid may lack a functional origin of replication for the host cell. Such non-replicability may

20 be conditional, e.g. replication may be temperature sensitive. The former is preferred, however, for simplicity of culturing and because incubation at high temperature can be mutagenic.

25 This feature assists in eliminating the plasmid from the host cell once a transposition event has occurred, especially if the cell is cultured under conditions selective for the first marker, but not the second marker.

30 Preferably the plasmid is pUC1169::pOJ260 or a derivative thereof, more preferably pKay1.

In a second aspect, the present invention provides a method of generating a mutation in a bacterial host cell, the method comprising:

5 (a) providing a bacterial donor cell comprising the plasmid of the first aspect; and

(b) introducing the plasmid from the donor cell to the host cell by conjugation, whereupon transposition of the transposable element into the host cell nucleic acid may occur.

10

The preferred features of this aspect and the previous aspect are interchangeable, except where the context requires otherwise.

15 The host cell is preferably a pre-germinated actinomycete spore, most preferably of the preferred genus, species and/or strains identified previously.

For transfer of (in particular) plasmids bearing an RP4 20 *oriT* to occur by conjugation, a transfer function should be supplied, preferably *in trans*, e.g. by an *E.coli* donor strain such as ET12567 carrying the self-transmissible pUB307 (Bennett *et al.*, 1977, Flett *et al.*, 1997) or ET12567 carrying the non-transmissible pUZ8002 (Kieser *et al.*, 25 2000).

Thus step (b) preferably includes such provision *in trans* 30 of transfer function. This step may involve transforming the plasmid into a donor strain carrying a non-transmissible transfer plasmid (e.g. ET12567/pUZ8002), followed by incubation under suitable conditions with the host cell (e.g. a pre-germinated actinomycete spore).

The conjugation step (b) may comprise an incubation step or steps to allow conjugation and subsequently transposition to occur. It is thought that transposition can occur under any physiological conditions, though a 5 temperature between about 30°C and 37°C is preferred for transposition.

Preferably the conjugation step (b) is followed by a selective incubation step (c) under conditions which 10 select for host cells in preference to donor cells (e.g. nalidixic acid to kill *E.coli* donor cells) and for the presence in the host cell of the first marker gene (e.g. exposure to an antibiotic to which resistance is provided by the first marker gene).

15 The selective incubation step may be followed by a step (d) of identifying host cells in which a transposition event has occurred; this is indicated by lack of the second marker gene in host cells which possess the first 20 marker gene. This may be performed by replica plating.

The mutant host cells produced according to this method may be stored for future use, in any suitable form, preferably (when the host cell is an actinomycete) as 25 spores.

The present invention also provides a host cell producible or as produced by the process of the second aspect. Furthermore, the invention provides a method of 30 determining the effect of a genetic disruption, the method comprising culturing such a host cell and determining the effect of the disruption on the cell.

In further aspects, the invention provides a donor cell comprising a plasmid according to the first aspect, and a kit comprising the plasmid, a donor cell and a transfer plasmid (e.g. as defined above) capable of providing

5 transfer function *in trans*. The donor cell of the kit may contain the plasmid of the first aspect and optionally also the transfer plasmid, or these components may be provided separately. The plasmid of the first aspect or kit containing such a plasmid may be packaged with

10 instructions for use in the method of the second aspect.

The skilled person is well aware of suitable techniques for the preparation of such cells, e.g. from Sambrook et al. 1989, and Kieser et al, 2000.

15

Preferably the method of the second aspect will be carried out simultaneously on several host cells, which following conjugation and transposition can be pooled to produce a library of independently mutated host cells.

20

Accordingly, the invention provides in a further aspect a method for producing a library of independently mutated host cells, the method comprising:

(a) providing a plurality of bacterial donor cells  
25 each comprising a plasmid according to of the first aspect;

(b) introducing plasmids from respective donor cells into respective host cells by conjugation, whereupon transposition of the transposable element into the host  
30 cell nucleic acid may occur; and  
pooling the host cells to form a library of host cells.

Again, preferred features of this aspect and the first two aspects are, unless the context requires otherwise, interchangeable. In particular, a selective incubation step (c) preferably occurs between step (b) and the 5 pooling step. However, the identification step (d) need not necessarily be carried out.

Preferably the library comprises at least 50 independent transposon mutant host cells, more preferably at least 80, 10 more preferably at least 100, 200, 500, 800, 1000, 2000, 5000, 8000 or 10000, still more preferably at least 20000, 50000 or 80000. "Independent" in this context means derived from independent transposition events. Thus cells (e.g. spores) cloned from a single mutant host cell are 15 not regarded as independent mutants. Different transposition events can result in identical mutants, however, and host cells having identical transposon mutations arising from different transposition events are regarded to be independent.

20

In a further aspect, the invention provides a library as produced or as producible according to the preceding aspect.

25 The library may be screened for altered phenotypic characteristics arising from the mutations in the host cells, e.g. for desired altered phenotypic characteristics.

30 Having identified a host cell displaying an altered phenotypic characteristic of interest, the gene or control sequence associated with the altered phenotypic characteristic may be identified by identifying the location of the transposable element in the mutated host

cell. The unmutated gene or control sequence may be cloned (e.g. from an unmutated host cell genomic library) using primers based on host cell genomic sequences flanking the transposable element in the mutated host cell.

The host cells may also be subjected to genomic DNA extraction procedures employed in the art, to provide a library of mutated genomic DNA.

10

#### Variants

References herein to genes and nucleic acids are not to be interpreted as being restricted to genes and nucleic acids having identical nucleic acid sequences to the specific genes and nucleic acids disclosed herein. Rather, genes and nucleic acids which are variants of those sequences are also included, provided that the functional interactions of the various components used in the methods and materials of the invention are maintained. Genes and nucleic acids having the specific sequences disclosed are preferred embodiments.

The term "variant" as used herein in relation to a particular nucleic acid (the reference nucleic acid) may denote: any nucleic acid having a sequence which is different from that of the reference nucleic acid, but which is its complement or which shows significant nucleic acid sequence identity with, or hybridisation under stringent conditions to, the reference nucleic acid or its complement or a fragment of the reference nucleic acid or its complement.

Similarly, a "variant" may be any nucleic acid having a sequence which encodes an amino acid sequence which is different from, but which shows significant amino acid sequence identity to, the amino acid sequence encoded by 5 the reference nucleic acid.

Significant nucleic acid sequence identity is preferably at least 50%, more preferably 60%, 70%, 80% or 90%, still more preferably 95%, 98% or 99%.  
10

Significant nucleic acid sequence identity is preferably shown between the variant nucleic acid (or a portion thereof) and a fragment of at least 30 residues of the reference nucleic acid, more preferably a fragment of a 15 least 60, 90 or 120 residues, still more preferably the entire reference nucleic acid.

Significant amino acid sequence identity, except where otherwise indicated, is preferably at least 70%, more 20 preferably 75%, 80%, 85%, 90%, 95%, 98% or 99%.

"Percent (%) nucleic acid sequence identity" is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues 25 in the sequence under comparison, as determined by the BLASTN module of WU BLAST-2 (Altschul et al. (1996); <http://blast.wustl.edu/blast/README.html>), set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.  
30

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M

sodium citrate/ 0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1%

5 polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution,

10 sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

15 "Percent (%) amino acid sequence identity" is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the sequence with which it is being compared, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. The % identity values used herein are generated by WU-BLAST-2 which was obtained from Altschul et al. (1996) Methods in Enzymology, 266:460-480; <http://blast.wustl.edu/blast/README.html>. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSPS and HSPS2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being

searched; however, the values may be adjusted to increase sensitivity. A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region, multiplied by 100. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WUBLAST-2 to maximize the alignment score are ignored).

10 Derivatives

A nucleic acid sequence (e.g. plasmid) which is referred to herein as being "derived" from another nucleic acid sequence has preferably actually been derived by manipulation of that other nucleic acid sequence (although it is contemplated that derivation may also involve de novo nucleic acid synthesis). This may be reflected by the presence of stretches of complete nucleic acid identity between the nucleic acid sequences, especially in intergenic sequences. Such identity in intergenic sequences is preferably shown over one or more stretch of at least 15, 20 or 25 nucleotides.

25 Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Fig. 1 shows the transposon Tn4556 and its viomycin-resistant derivative Tn4560 (from Kieser et al., 2000; page 578). Tn4556 (GenBank accession number M29297) is a 6.8 kb transposable element from *Streptomyces fradiae*. Since Tn4556 contains no known markers for resistance to antibiotics or heavy metals, a 2 kb *BamHI* fragment of a promoterless viomycin resistance gene (*vph*) was cloned

into Tn4556. This viomycin-resistant derivative of Tn4556 was termed Tn4560, and construction of Tn4560 was fully described by Chung (1987).

5 Key: IR - 38 bp imperfect terminal inverted repeat; *tnpA* - transposase; *tnpR* - resolvase; *res* - resolution site; *vph* - viomycin phosphotransferase gene from *Streptomyces vinaceus*, which encodes viomycin resistance (GenBank accession number X02393).

10

**Fig. 2** shows plasmid pUC1169 (from Kieser et al., 2000; page 579). Tn4560 was cloned into pMT660 to generate pUC1169. The construction of pUC1169 was fully described by Chung (1987). pMT660 (Birch and Cullum, 1985) is a 15 temperature-sensitive derivative of pIJ702 (Katz et al., 1983), itself a thiostrepton-resistant derivative of pIJ101 (Kieser et al., 1982; GenBank accession number M21778). Features within the shaded region form part of Tn4560 and are described in Figure 1.

20

Key: *rep* pIJ101 - gene encoding the replication protein from pIJ101; *ori* pIJ101 - origin of replication for *Streptomyces*, from pIJ101; *tsr* - 23S A<sup>1067</sup> rRNA methylase from *Streptomyces azureus*, which encodes thiostrepton 25 resistance (GenBank accession numbers X02392, X54219).

Maps of pIJ702, pIJ101, and *tsr* can be found in Kieser et al. (2000), pages 536, 254, and 456, respectively. The unique *BamHI* site used for cloning of pOJ260 during 30 construction of pUC1169::pOJ260 is indicated at 12 o'clock on the map (downward arrow,  $\downarrow$ ). The unique *BstBI* site used for introduction of a frameshift and STOP codon into the *rep* gene of pIJ101 in pUC1169::pOJ260, during the construction of pKay1, is also indicated (symbol

consisting of a white cross on a black diamond-shaped background, i.e. ♦).

Fig. 3 shows plasmid pOJ260 (from Kieser et al., 2000; page 566). pOJ260 is a suicide vector for *Streptomyces* containing *oriT* RK2 for conjugation from *E.coli* to *Streptomyces*. Construction of pOJ260 was fully described by Bierman et al. (1992).

10 Key: *ori* pUC18 - *E.coli* origin of replication from pUC18 (Yanisch-Perron et al., 1985); *oriT* RK2 - origin of transfer from the IncP-group plasmid RK2 (Guiney and Yakobson, 1983); *aac(3)IV* - amino glycoside acetyltransferase from *Klebsiella*, which encodes apramycin resistance (GenBank accession number X99313).

A map of the *aac(3)IV* fragment can be found in Kieser et al. (2000), page 465.

20 The unique *BamHI* site in *lacZα* used for cloning of pOJ260 into pUC1169 during construction of pUC1169::pOJ260 is indicated at 12 o'clock on the map (downward arrow). This allowed blue-white screening of recombinants.

25 Fig. 4 shows plasmids pUC1169::pOJ260 and pKay1.

(a) pUC1169 (Chung, 1987) and pOJ260 (Bierman et al., 1992) were digested with *BamHI* and ligated, to generate pUC1169::pOJ260.

30 (b) pKay1 was generated by introducing a 2 bp frameshift and STOP codon at the unique *BstBI* site of pUC1169::pOJ260 (symbol consisting of a white cross on a black diamond-shaped background). pUC1169::pOJ260 was digested with

*BstBI* (TT<sub>1</sub>CGAA), and the 2 bp 5' overhang was repaired with the Klenow fragment. Blunt pUC1169::pOJ260 was re-ligated with T4 DNA ligase, to generate pKay1. The *BstBI* site was replaced with a *NruI* site (TTCG<sub>1</sub>CGAA).

5

**Fig. 5** shows the construction of a Tn4560 mutant "megalibrary". 96 conjugations between *E.coli* ET12567 (pUZ8002, pKay1) and *S.coelicolor* M145 were performed. Each conjugation yielded ~1000 independent mutants. Spores from each conjugation were pooled to form 96 "minilibraries". Spores from ten minilibraries were pooled to form a "maxilibrary" (containing ~10,000 independent mutants). Ten maxilibraries were generated in this way. Pooling equal volumes of each maxilibrary generated a "megalibrary", containing ~100,000 independent transposon mutants. Each minilibrary, maxilibrary and the megalibrary consists of spores and the corresponding genomic DNA.

20 **Fig. 6** shows screening of the Tn4560 mutant library for insertions in a cytochrome P450 gene.

(a) The arrangement of genes on *S. coelicolor* cosmid SC7E4, with the co-ordinates of each gene within the 25 cosmid and the predicted size of each gene. The position of the gene-specific primer, "SC7E4.19", is indicated, immediately upstream of the beginning of the sesquiterpene cyclase gene.

30 (b) The location of Tn4560-specific primers. Tn4560 can insert in either orientation into the chromosome of *S. coelicolor* M145, as indicated by (+) and (-). Use of distinct primers for each end of Tn4560 enables

determination of Tn4560 orientation. The distance of each primer from the end of Tn4560 is indicated.

To identify insertions in the cytochrome P450 gene, PCR  
5 products ranging between 1400 bp and 2790 bp with  
"SC7E4.19" and "NEST4", and between 1280 bp and 2670 bp  
with "SC7E4.19" and "TN4560R-2", were expected. A PCR  
product of 1.7 kb observed with "SC7E4.19" and "TN4560R-2"  
was predicted to represent a Tn4560 insertion in the  
10 cytochrome P450 gene.

**Example 1****Construction of pUC1169::pOJ260**

A fusion of pUC1169::pOJ260 was constructed to enable  
5 mobilisation of Tn4560 from *E. coli* to *S. coelicolor* M145.  
pOJ260 (Bierman et al., 1992) contains the 760 bp *oriT*  
fragment from the IncP-group plasmid RK2 (Guiney and  
Yakobson, 1983), an *E. coli* origin of replication (*ori*)  
from pUC (Yanisch-Perron et al., 1985), and an apramycin  
10 resistance marker (*aac(3)IV*) (see Figure 3). pUC1169 and  
pOJ260 both contain a unique *Bam*HI site. In pOJ260, the  
*Bam*HI site is in the multiple cloning site (MCS) in *lacZα*  
(enabling blue-white screening of recombinant clones); in  
pUC1169, the *Bam*HI site is near the 3' end of the *rep* gene  
15 from pIJ101. Cloning of pOJ260 at this site would be  
predicted to disrupt the *rep* gene and generate a construct  
that is non-replicative in *Streptomyces*.

pUC1169 and pOJ260 were digested with *Bam*HI (see Figures 2  
20 and 3 respectively for locations of *Bam*HI sites),  
according to standard protocols (Ausubel et al., 1989;  
page 3.1.1). Digestions were checked by agarose gel  
electrophoresis - *Bam*HI linearises pUC1169 and pOJ260, to  
give single fragments of 14.3 kb and 3.5 kb respectively.  
25 *Bam*HI digested pUC1169 and pOJ260 were purified prior to  
ligation by adjusting the digest volume up to 100μl and  
extraction with an equal volume (100μl) of  
phenol:chloroform 1:1, removing the upper aqueous layer  
and subsequently extracting with an equal volume of  
30 chloroform. DNA was then precipitated by the addition of  
0.1 volumes of 3M sodium acetate (unbuffered) to the  
aqueous layer followed by 1 volume of isopropanol. The  
ethanol mixture was left at -20°C for 1 hour before  
centrifugation in a microfuge at 13,000 rpm for 5 minutes.

The DNA pellets were washed with 70% ethanol and air dried before resuspending in 20 $\mu$ l TE pH8.0 (10mM Tris-HCl, 1mM EDTA). *Bam*HI-digested pUC1169 and pOJ260 were ligated (1 $\mu$ l of each out of the 20 $\mu$ l above) in a 1:1 ratio with 5 unit of T4 DNA ligase in a total volume of 10  $\mu$ l, according to standard protocols (Ausubel et al., 1989; page 3.16.1), except that ATP was not added to the ligation mixture (the 10x T4 DNA ligase buffer provided by the manufacturer (Roche) already contains 10mM ATP).

10 Ligation was performed overnight in a 12°C waterbath.

Competent cells of *E. coli* DH5 $\alpha$  were prepared as follows. A single colony of *E. coli* DH5 $\alpha$  was inoculated into 10 ml Lennox broth (LB; Kieser et al., 2000; page 412) and grown 15 overnight at 37°C with vigorous shaking (300 rpm in a rotary shaker). 0.5 ml of the overnight culture was inoculated into 50 ml pre-warmed LB and grown at 37°C until an OD<sub>600nm</sub> of 0.5–0.6 was reached. The culture was transferred to a sterile, disposable, ice-cold 50 ml 20 polypropylene tube, and chilled on ice for 10 min. The cells were pelleted by centrifugation at 2500 rpm and 4°C for 10 min. The cell pellet was resuspended in 15 ml ice-cold TfblI buffer (30mM KCOOCH<sub>3</sub>, 100mM RbCl, 10mM CaCl<sub>2</sub>.2H<sub>2</sub>O, 50mM MnCl<sub>2</sub>.4H<sub>2</sub>O, 15% (w/v) glycerol, adjusted to 25 pH 5.8 with 0.2M acetic acid, filter sterilised, stored at 4°C), and left on ice for 20 min. Centrifugation was repeated, the cell pellet was resuspended in 2 ml ice-cold TfblII buffer (10mM PIPES, 10mM RbCl, &5mM CaCl<sub>2</sub>.2H<sub>2</sub>O, 15% (w/v) glycerol, adjusted to pH 6.5 with 1M KOH, filter 30 sterilised, stored at 4°C), and left on ice for 10 min. Competent cells were dispensed as 100  $\mu$ l aliquots into Eppendorf tubes, and were either used immediately for

transformation, or were frozen rapidly in a dry ice/ethanol bath, before storage at -70°C.

For transformation, a 100 µl aliquot of *E. coli* DH5 $\alpha$

5 competent cells was thawed slowly on ice, half (5 µl) of the pUC1169::pOJ260 ligation mixture was added and mixed by gently flicking the bottom of the tube, and the mixture incubated on ice for 30 min. The cells were heated for 90 s in a 42°C waterbath, and immediately chilled on ice for 2

10 min. 1 ml LB was added, and the cells were incubated for 90 min in a 37°C waterbath, mixing occasionally. 100 µl (per plate) of transformed competent cells were spread onto L agar (Kieser et al., 2000; page 411), containing isopropyl- $\beta$ -D-thiogalactoside (IPTG), 5-bromo-4-chloro-3-

15 indoly- $\beta$ -D-galactoside (X-gal) and apramycin (Am) - per 200 ml L agar, 200 µl of each of 0.1M IPTG, 40 mg/ml X-gal, and 50 mg/ml Am (Am50; 50 µg/ml final) were added. Following growth of colonies after incubation at 37°C ten white apramycin-resistant (Am<sup>R</sup>) colonies were inoculated

20 into 5 ml LB containing 5 µl of Am50, and grown overnight at 37°C with vigorous shaking. Plasmid DNA was isolated (Sambrook et al., 1989; page 1.25) and digested with BamHI using standard conditions (Sambrook et al., 1989). Digests were checked by agarose gel electrophoresis. One

25 plasmid preparation gave the expected BamHI restriction pattern - two bands of 14.3 kb and 3.5 kb, corresponding to linearised pUC1169 and pOJ260 respectively (see Figure 4(a) for a plasmid map of pUC1169::pOJ260).

30 For mobilisation into *S. coelicolor* M145, pUC1169::pOJ260 was transformed into competent cells of *E. coli* ET12567 (pUZ8002), (obtained from the John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH,

UK). Preparation and transformation of competent cells was as described above for *E. coli* DH5 $\alpha$ . ET12567 is a methylation-deficient *E. coli* strain (*dam13::Tn9 dcm6 hsdM hsdR recF143 zjj201::Tn10 galK2 galT22 ara14 lacY1 xy15 5 leuB6 thiI tonA31 rpsL136 hisG4 tsx78 mtII glnV44 F-*; MacNeil et al., 1992) and must be used as the donor strain when introducing DNA into *S. coelicolor* M145, which possesses a methyl DNA-specific restriction system.

pUZ8002 (Kieser et al 2000) is a non-transmissible plasmid 10 that is thought to supply the transfer functions in *trans*. Transformants were selected by growth at 37°C on L agar plates containing Am50. Single Am<sup>R</sup> colonies were inoculated into 10 ml LB containing 10  $\mu$ l of each of Am50, 15 25 mg/ml chloramphenicol (Cml25; 25  $\mu$ g/ml final; to select for *dam13::Tn9*), and 25 mg/ml kanamycin (Km25; 25  $\mu$ g/ml final; to select for pUZ8002). Plasmid DNA was isolated from liquid cultures and digested with *Bcl*I using standard conditions (Sambrook et al., 1989). This confirmed the 20 non-methylated status of the plasmid DNA, and the expected restriction pattern of 1005/1055, 1377, 1772, 3042, 4367 and 5192 bp fragments was observed (*Bcl*I is blocked by *dam* methylation).

Conjugal transfer of pUC1169::pOJ260 from *E. coli* to *S. coelicolor* M145 25

The protocol used was as described by Kieser et al. (2000; page 249). A single colony of *E. coli* ET12567(pUZ8002, pUC1169::pOJ260) was inoculated into 10 ml LB containing 10  $\mu$ l of each of Am50, Cml25 and Km25, and grown overnight 30 at 37°C with vigorous shaking. The overnight culture was diluted 1:100 in 5 ml LB plus 5  $\mu$ l of each of Am50, Cml25 and Km25, and grown at 37°C with vigorous shaking to an OD<sub>600nm</sub> of 0.4-0.6. The cells were washed twice with 5 ml

LB (to remove antibiotics) with centrifugation after each wash in a bench top centrifuge at 3000 rpm for 5 minutes to pellet the cells and finally resuspended in 0.5 ml LB. While washing the *E. coli* cells,  $10^8$  *S. coelicolor* M145 spores were added to 0.5 ml 2x YT (Kieser et al., 2000; page 414), heat shocked in a 50°C waterbath for 10 min (to induce germination), and cooled to room temperature (RT). 0.5 ml *E. coli* cells (ET12567 prepared as described above) were combined with 0.5 ml heat-shocked spores, and mixed by vortexing. The cell mixture was centrifuged briefly at 13,000 rpm in a microcentrifuge, the supernatant discarded, and the cell pellet resuspended in the residual liquid (~100µl). The suspension was spread onto one dry plate of mannitol soya flour (MS) agar (Kieser et al., 2000; page 409) supplemented with 10mM MgCl<sub>2</sub> (per 100 ml MS agar, 1 ml of 1M MgCl<sub>2</sub> was added). After incubation at 30°C for 16-20 h, the plate was flooded with 0.5 ml sterile distilled water containing 25 µl of each of 20 mg/ml nalidixic acid (Ndx20; 20 µg/ml final) to kill the *E. coli* donor, and 30 mg/ml viomycin (Vio30; 30 µg/ml final). The plate was left in a laminar flow cabinet until all the liquid had been absorbed, and was then incubated at 30°C for 3-4 days, after which time growth of viomycin-resistant (Vio<sup>R</sup>) *S. coelicolor* M145 exconjugants was very dense, almost confluent (~10,000 colonies per plate).

#### Testing of pUC1169::pOJ260

Spores from 22 distinct areas of the plate were streaked for single colonies on MS agar containing Vio30 and Ndx20 (per 100 ml MS agar, 100 µl of each of Vio30 and Ndx20 were added). A single colony from each streak was spread over a 1 cm<sup>2</sup> area (patched using sterile wooden toothpicks) on MS agar plates containing Vio30 and Ndx20, with several

patches per plate in a grid pattern. Patches were replicated to MS agar containing thiostrepton or Am50 (per 100 ml MS agar, 100 µl of each of 50 mg/ml thiostrepton (Thio50; 50 µg/ml final) and Am50). pUC1169::pOJ260 5 should not replicate in *S. coelicolor* M145, because of disruption of the rep gene. It was expected that Vio<sup>R</sup> resulted from transposition of Tn4560 into the chromosome, and that the exconjugants would be apramycin-sensitive (Am<sup>S</sup>) and thiostrepton-sensitive (Thio<sup>S</sup>). However, all 22 10 patches were Am<sup>S</sup>, but unexpectedly 14/22 (64%) were thiostrepton-resistant (Thio<sup>R</sup>). It appears that pUC1169::pOJ260 may be unstable when transferred to *S. coelicolor* M145, in some cases generating an intact, replicative pUC1169 molecule, with loss of pOJ260 (which 15 is non-replicative in *Streptomyces*). This would give rise to Thio<sup>R</sup> Am<sup>S</sup>. Hence the actual generation rate of Vio<sup>R</sup> only exconjugants was 36%, considerably lower than would be expected for an efficient *in vivo* random transposon mutagenesis method.

20

Spores were scraped from the surface of 15 patches and inoculated into 10 ml TSB:S-YEME liquid cultures containing 10 µl of each of Vio30 and Ndx20. TSB:S-YEME is a 1:1 (v:v) mixture of tryptone soya broth (TSB; Kieser 25 et al., 2000; page 412) and yeast extract-malt extract medium (YEME) supplemented with 20% glycine (0.5% final; Kieser et al., 2000; page 412). Cultures were grown at 30°C for 40 h, after which time the mycelia were harvested by centrifugation at 3000 rpm in a bench top centrifuge, 30 and washed with an equal volume of sterile distilled water. The centrifugation step was repeated and the mycelia were washed a second time with an equal volume of sterile distilled water. Following centrifugation the supernatant was removed leaving the mycelial pellet and

residual liquid. Washed mycelia (~1 µl of the pellet and residual liquid) was used as a PCR template to confirm the status of *vph* and *tsr* markers. Positive (10 ng pUC1169 DNA) and negative (*S. coelicolor* M145 mycelia) controls 5 were also included. A multiblock Stratagene Robocycler™ was programmed as follows: 1 cycle of 95°C 10 min; 30 cycles of 95°C 1 min, 55°C 2 min, 72°C 1 min; 1 cycle of 72°C 10 min. PCR was performed using 0.2 ml thin-walled PCR tubes. The PCR mixture contained 5 µl 10x PCR buffer 10 (1x final; Applied Biosystems), 5 µl dimethyl sulphoxide (DMSO; 5% final), 0.4 µl 25mM dNTP mixture (200µM final), 1 µl (50 pmol/µl) of each of "TSR1" and "TSR2" (for amplification of *tsr*), or "VPH1" and "VPH2" (for amplification of *vph*) ("TSR1" - 5'-TCGACTCCTCGATCGTCAACC- 15 3'; "TSR2" - 5'-GAGCGTCATCAGCTGCATACC-3'; "VPH1" - 5'-CACCGGGTGGTCTGCTTC-3'; "VPH2" - 5'-GTGGCGAACTCCTGCCAC-3'), and 1 unit AmpliTaq Gold® DNA polymerase (Applied Biosystems), in a total volume of 50 µl. All 15 Vio<sup>R</sup> exconjugants gave a PCR product equivalent to the *vph* gene 20 (~350 bp), whereas 8 gave a PCR product equivalent to the *tsr* gene (~350 bp). These 8 exconjugants corresponded to Thio<sup>R</sup> patches. This confirmed that Thio<sup>R</sup> was due to the presence of *tsr*, and was not spontaneous resistance.

**Example 2****Construction of pKay1**

Introduction of a frameshift (and a STOP codon) in the  
5 pIJ101 *rep* gene of pUC1169::pOJ260 should render any  
pUC1169-related molecule non-replicative in *Streptomyces*.  
This should enable efficient *in vivo* generation of a  
library of *Streptomyces* mutants by transposon mutagenesis.  
pUC1169::pOJ260 contains a unique *Bst*BI site in the *rep*  
10 gene of pIJ101. Digestion of DNA by *Bst*BI should generate  
a 2 bp 5' overhang, which can be repaired to generate  
blunt ends. Re-ligation of the blunt ends of  
pUC1169::pOJ260 should eliminate the *Bst*BI site and  
introduce a 2 bp frameshift and a STOP codon. In place of  
15 the *Bst*BI site, a *Nru*I site should also be introduced  
(Figure 4(b)).

pUC1169::pOJ260 was digested with *Bst*BI in a total volume  
of 20 $\mu$ l. The 2 bp 5' overhang was repaired using the  
20 Klenow fragment (by the addition of Klenow fragment and  
dNTP's to the 20 $\mu$ l *Bst*BI digest to give a final volume of  
23 $\mu$ l), according to standard protocols (Ausubel et al.,  
1989; page 3.5.8). Blunt ends were re-ligated using 5  
units of T4 DNA ligase (using 4 $\mu$ l of the 23 $\mu$ l blunting  
25 reaction above in a total final volume of 20 $\mu$ l), as  
previously described. Competent cells of *E. coli*  
ET12567 (pUZ8002) were transformed with quarter (5  $\mu$ l) of  
the ligation mixture, as previously described, and  
transformants were selected on L agar containing Am50,  
30 Cml25 and Km25. Resistant colonies were inoculated into  
10 ml LB containing 10  $\mu$ l of each of Am50, Cml25, and  
Km25, and plasmid DNA was prepared using standard  
conditions (Sambrook et al., 1989). Plasmid DNA was

digested with *Bst*BI, *Nru*I, and *Bcl*I. If in-filling of the *Bst*BI site had been successful, to generate a new *Nru*I site, the following pattern of *Nru*I restriction fragments would be observed - 199, 259, 597, 1149, 2116, 6190 and 5 7300 bp. This pattern was observed for 3/6 samples. These 3 samples did not digest with *Bst*BI, confirming loss of the *Bst*BI site. Plasmid DNA was non-methylated, as confirmed by *Bcl*I digestion. The correct construct had been made, and this was called pKay1.

10

**Testing pKay1**

pKay1 was transferred by conjugation from *E. coli* ET12567 (pUZ8002, pKay1) to *S. coelicolor* M145, as previously described. Spores from twenty distinct areas 15 of the plate were streaked for single colonies on MS agar containing Vio30 and Ndx20. A single colony from each streak was patched on MS agar plates containing Vio30 and Ndx20, as previously described. Patches were replicated to Difco Nutrient Agar (DNA; Kieser et al., 2000; page 20 411) plates containing Vio30 and either Thio50 or Am50. All twenty Vio<sup>R</sup> patches were also Thio<sup>S</sup> Am<sup>S</sup>, (i.e. 100%) indicating that pKay1 had been lost, and that Tn4560 is present in the chromosome. Stability of pKay1 when transferred to *S. coelicolor* M145 can not be determined, 25 because pKay1 appears to have effectively operated (as intended) as a suicide plasmid.

Spores were scraped from the surface of each patch, and used to inoculate 10 ml TSB:S-YEME liquid cultures 30 containing 10 µl of each of Vio30 and Ndx20. Cultures were grown at 30°C for 40 h, after which time the mycelia were harvested and washed as previously described. Washed mycelia (~1 µl) were used as a PCR template to confirm presence of the *vph* marker and loss of the *tsr* marker (and

hence loss of pKay1), using primers "TSR1" and "TSR2" for the *tsr* gene, and primers "VPH1" and "VPH2" for the *vph* gene, as previously described. All twenty *Vio*<sup>R</sup> exconjugants gave a PCR product corresponding to the *vph* gene (~350 bp), whereas none gave a PCR product corresponding to the *tsr* gene (~350 bp).

**Construction of the Tn4560 mutant "Megalibrary"**

96 independent conjugations between *E. coli* 10 ET12567 (pUZ8002, pKay1) and *S. coelicolor* M145 were performed, as previously described. Conjugation plates were incubated at 30°C for approximately 7 days to allow profuse sporulation. Spores (representing ~1000 independent mutants) from a single conjugation plate were 15 then pooled to form a "minilibRARY" (as described in Kieser et al., 2000; page 169) and resuspended in 1ml 20% glycerol for frozen storage. This was performed for each conjugation plate resulting in a total of 96 "minilibraries". Spores from ten minilibraries were pooled 20 in equal 100µl volumes to form a "maxilibRARY" (containing ~10,000 independent mutants). Ten maxilibraries were generated in this way (referred to as 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-86, and 87-96). Maxilibrary 81-86 contains spores from only six 25 minilibraries. Pooling equal volumes of each maxilibrary generated the "Megalibrary" (1-96), containing ~100,000 independent transposon mutants. Each minilibrary, maxilibrary and the megalibrary consists of spores and the corresponding genomic DNA. For the preparation of DNA, 10-30 20 µl of the spore suspension was used to inoculate 10 ml TSB:S-YEME liquid medium containing Vio30 and Ndx20 to eliminate non-resistant strains and also to inhibit residual *E. coli* cells from the conjugation experiment. Genomic DNA was isolated from cultures using the salting

out procedure of Pospiech and Neumann (1995) as described in Kieser et al., (2000); page 169.

**Mutant representation in the library pools**

5 When exconjugant colonies on MS agar plates were only just visible to the eye, these small single colonies were picked to fresh MS agar plates containing Vio30 and Ndx20. Only 10% of the colonies were observed to survive. Given that a confluent plate represents approximately 10,000

10 colonies, this would suggest the presence of a minimum of ~1000 genuine Vio<sup>R</sup> colonies per conjugation plate. However, it is unlikely that all of these small colonies continue to grow and sporulate in the presence of Vio30 and will not therefore be represented in the final pool of spores

15 generated from a confluent plate. The actual number of genuine Vio<sup>R</sup> colonies per plate is therefore likely to be much higher than 1000 and this figure can be considered to be a worst case scenario. Approximately  $5 \times 10^8$  spores are obtained per plate (see Titration of Spores in later

20 section) and assuming the worst case scenario that all the non-viable mini colonies described above do continue to grow and produce spores (and only 10% are genuine Vio<sup>R</sup> transposon mutants), and that each colony produces the same number of spores, each mutant type (~1000 types)

25 would be represented by a minimum of  $5 \times 10^4$  spores on average in the entire 1 ml "minilibrary" spore suspension, giving a >99% chance that all mutants are represented in the sample. It appears that fewer colonies and therefore spores are produced by the non-viable mini-colonies than

30 by the genuine Vio<sup>R</sup> mutants, making the samples even more representative.

The DNA isolations (made from Vio<sup>R</sup> cultures) produced ~ 1 mg genomic DNA for each library and approximately 100 ng

samples are used for PCR amplification. The *S. coelicolor* genome is ~8000kb, giving approximately  $10^7$  transposon containing genomes per ng of DNA; again making these samples highly statistically representative. Mutant types 5 that grow poorly in submerged culture are, however, lost.

**Screening of the Tn4560 mutant library for insertions in a specific gene**

The genes of interest, on cosmid SC7E4 (GenBank accession 10 number AL359214; cosmid available from Dr H.M Kieser, John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK), were predicted to encode a sesquiterpene cyclase (SC7E4.19, 17196..18281) and a putative cytochrome P450 (SC7E4.20, 18278..19663) (see 15 Figure 6(a)). A Tn4560 insertion mutant in the sesquiterpene cyclase had previously been identified during sequencing of randomly picked Vio<sup>R</sup> *S. coelicolor* M145 exconjugants. In order to isolate a mutant in the downstream putative cytochrome P450 gene, the Tn4560 20 mutant library was first screened for insertions in the cytochrome P450, using PCR.

A gene-specific primer was designed upstream of SC7E4.19 ("SC7E4.19": 5'-GCGGCTGAGAGGGGTTGTTTC-3'), and was used in 25 combination with a primer specific for the left end of Tn4560 ("NEST4": 5'-AGGAACACGCCGAGATAGCG-3') and, in a separate reaction, with a primer specific for the right end of Tn4560 ("TN4560R-2": 5'-CAGAATTCCCCCTTGCCACAGATAACAG-3') (see Figure 6(b)). PCR 30 conditions were as described previously. PCR was performed for each maxilibrary, using 100 ng of the corresponding genomic DNA as template, and 50 pmoles of each primer. 12 PCR products were observed by agarose gel electrophoresis (6 with "NEST4", 6 with "TN4560R-2"),

ranging in size from 600 bp to 2 kb. One PCR product of ~1.7 kb, obtained with "TN4560R-2" and maxilibrary 1-10, was selected for further investigations. This PCR product should, on the basis of size, represent an insertion of 5 Tn4560 in SC7E4.20 (see Figure 6(b)). PCR was repeated using genomic DNA from minilibraries 1 to 10 as template. The PCR product of interest was observed in minilibrary 3. The PCR product was purified using the QIAquick™ PCR Purification Kit (QIAGEN), according to manufacturer's 10 instructions, except that the eluate was in 30 µl distilled water, rather than the elution buffer supplied with the kit. 5 µl of the eluate was checked by agarose gel electrophoresis. The DNA concentration (in µg/µl) was determined spectrophotometrically at OD<sub>260</sub> and OD<sub>240</sub>, using 15 a Pharmacia GeneQuant II spectrophotometer.

**ABI sequencing of PCR product and analysis of returned sequence**

The sequencing PCR was done in the Stratagene Robocycler™. 20 Conditions for the sequencing PCR reaction were as follows: 30 cycles of 95°C 30 s, 50°C 30 s, 60°C 4 min. The sequencing reaction contained 200 ng PCR product (taken directly from the eluate), 1 µl "TN4560R-2" (1 pmol/µl), 1 µl DMSO (5% final), 4 µl ABI PRISM® BigDye™ 25 Terminators v3.0 cycle sequencing kit (Applied Biosystems), and made up to 10 µl with distilled water. Once the sequencing PCR was complete, the reaction mixture was transferred to a 0.5 ml Eppendorf tube. 3.5 µl 7.5M ammonium acetate and 34 µl ethanol were added, and the 30 tube was placed in the freezer for 1 h. The tube was centrifuged at 13,000 rpm for 20 min in a microcentrifuge. The supernatant was carefully poured off, the invisible pellet was washed in 70% ethanol, and allowed to air dry.

Dried samples were then sequenced following standard procedures at the John Innes Centre central sequencing facility (Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK). The sequence obtained is shown below.

5 Nucleotides in lower case represent the sequence of Tn4560 (1..155) while nucleotides in upper case represent *S. coelicolor* genomic DNA (156..387):

10 1 gcgtcaccgc aggtgacgga tcaagctacc gcagatattg gggtgttatac  
15 51 tgccggaaaga cttcagacga gtccgcccgc cgacgctccg tcagccgccc  
20 101 acctgggccc actgatcatt agcgccggtt ttcgttcgga tggccctcaa  
25 151 cccccTCCTC CTCCATGATC GGCCCGTAGG CGGGGATGGC GTCTAGCCGG  
30 201 AACGCCGGCT GGATGGTGCG CCGCTGGCGG CGGTGGAGCG GGCGTGGC  
35 251 GGTGCCACG CCCTCCTTGC CGAGCAGGCC CTCCAGCGAC TCCCACAGCG  
40 301 GACCGGCGAT GTGGTAGTCG GGGTTCAAGGG CGAGGGCGCC GGTCAAGCTCG  
45 351 GGGTCGTGA CGGCGTACAC GGTCTTCGGG CCGAGCT  
50 25 The genomic sequence (upper case) was put into a BLASTN (nucleotide-nucleotide) search against *S. coelicolor* cosmid sequences at the Sanger Centre website ([http://www.sanger.ac.uk/Projects/S\\_coelicolor/blast-server.shtml](http://www.sanger.ac.uk/Projects/S_coelicolor/blast-server.shtml)). The search parameters were the default options with "Filter low complexity regions" selected.  
30 The results of the BLASTN search confirmed that the Tn4560 insertion was in the putative cytochrome P450 gene (SC7E4.20), and gave us the precise Tn4560 insertion site within this gene (the co-ordinates of the 5bp direct

repeat sequence generated by Tn4560 insertion are 18686..18690).

**Design of oligonucleotide probe for colony hybridisation**

5 In order to isolate the specific mutant strain from the library of mutants, with Tn4560 inserted into SC7E4.20, colony hybridisation was used. For this, an oligonucleotide probe was designed, complementary to the above sequence. This 20 nt probe consists of 10 nt complementary to the end of Tn4560 and 10 nt complementary to the *S. coelicolor* M145 chromosome at the precise point of Tn4560 insertion ("P450::TN4560": 5'-TGGAGGAGGGAGGGGTTGAG-3').

15 **Preparation of the  $\gamma^{32}\text{P}$ -labelled oligonucleotide probe**  
3  $\mu\text{l}$  "P450::TN4560" (5 pmol/ $\mu\text{l}$ ), 10  $\mu\text{l}$   $\gamma^{32}\text{P}$  ATP, 2  $\mu\text{l}$  10 $\times$  T4 polynucleotide kinase (PNK) buffer (1 $\times$  final), 1  $\mu\text{l}$  T4 PNK, and 4  $\mu\text{l}$  distilled water were mixed in a screw capped Eppendorf tube, and incubated at 37°C for 30 min. T4 PNK  
20 was inactivated by incubating the tube at 70°C for 10 min. 1  $\mu\text{l}$  glycogen, 2  $\mu\text{l}$  7.5M ammonium acetate and 58  $\mu\text{l}$  ethanol were added, and the tube was placed on ice for 30 min. The tube was centrifuged at 13,000 rpm in a microcentrifuge for 20 min. The supernatant was  
25 discarded, and the probe was resuspended in hybridisation buffer. The tube was rinsed out with hybridisation buffer to ensure any probe stuck to the inside of the tube was removed. 200 ml hybridisation buffer was prepared according to the Boehringer Mannheim DIG manual - 3 g  
30 blocking powder was dissolved at 70°C for 1 h in 50 ml 20 $\times$  SSC, 2 ml 10% (w/v) N-lauroylsarcosine, 400  $\mu\text{l}$  10% (w/v) SDS, and 147.6 ml distilled water.

**Titration of spores**

In order to determine the spore titre for minilibrary 3, a 10-fold serial dilution series was set up, from  $10^0$  to  $10^{-8}$ . 100  $\mu$ l of each dilution was spread on MS agar plates containing Vio30. After 3-4 days incubation at 30°C, colonies were counted, and a titre of  $4.46 \times 10^8$  spores/ml was calculated.

**Preparation of filters**

10 Ideally, to obtain 99% coverage of the 1000 independent Tn4560 mutants in minilibrary 3, 5000 colonies should be plated out, assuming that each mutant is equally represented within the spore pool. 100  $\mu$ l of  $1 \times 10^{-5}$  and  $2 \times 10^{-6}$  dilutions of spores from minilibrary 3 were spread 15 on MS agar plates containing Vio30 and Ndx20 (two plates of the  $1 \times 10^{-5}$  dilution, and four plates of the  $2 \times 10^{-6}$  dilution) and incubated at 30°C for 3-4 days. There were approximately 784 and 780 colonies on the  $1 \times 10^{-5}$  plates, and 177, 160, 132, and 177 colonies on the  $2 \times 10^{-6}$  plates, 20 giving a total of 2210 colonies, and an 89% probability that the correct mutant would be found at least once.

Colonies were replicated in duplicate to nylon discs (OSMONICS INC., MAGNA nylon transfer membrane, gridded, 25 0.45  $\mu$ , 82 mm, Cat. No. N00HG08250) placed on the surface of L agar plates containing Ndx20. Colonies were also replicated to L agar plates containing Ndx20 without nylon discs. The replicates were incubated overnight at 30°C, and then left at room temperature (RT) for a second night.

30

At each stage during the preparation of the filters, the nylon membranes were transferred using a pair of broad forceps to sheets of Whatman paper in the base of large

square Petri dishes ( $22\text{ cm}^2$ ). A single  $22\text{ cm}^2$  sheet of Whatman paper could absorb 30 ml of solution. Filters were always transferred colony side up, and care was taken to transfer the filters in a horizontal position. Between 5 each treatment, filters were transferred to dry sheets of Whatman paper, to remove excess liquid, before proceeding to the next step. The filters were removed from the plates, and incubated at  $37^\circ\text{C}$  for 20 min with lysozyme solution (2 mg/ml in P buffer [Kieser et al., 2000; page 10 415]), to allow formation of protoplasts. The filters were incubated at RT for 7 min with denaturation solution (0.5M NaOH, 1.5M NaCl). The filters were incubated twice at RT for 5 min with neutralisation solution (3M NaCl, 0.5M Tris-HCl, pH 7.5), and allowed to air dry. DNA was 15 fixed to the filters by baking at  $80^\circ\text{C}$  for 1 h.

#### Hybridisation

Filters were arranged in hybridisation tubes so that they did not overlap, colony side facing towards the centre of 20 the tube. Filters were pre-hybridised at  $65^\circ\text{C}$  for 1 h in pre-warmed hybridisation buffer. Hybridisation buffer was discarded, and pre-warmed  $\gamma^{32}\text{P}$ -labelled "P450::TN4560" probe in hybridisation solution (~20 ml) was added. The filters were hybridised overnight at  $55^\circ\text{C}$ .

25

#### Washing and detection

The probe solution was decanted. The filters were washed four times with 5x SSC 0.1% SDS at  $55^\circ\text{C}$ , followed by four washes with 2x SSC 0.1% SDS at  $55^\circ\text{C}$ . The filters were 30 exposed to a phosphorimager screen overnight, and also to Kodak Biomax MS x-ray film for 5 h at  $-70^\circ\text{C}$ . A distinct strongly hybridising spot was observed on one plate from

the  $2 \times 10^{-6}$  dilution, which corresponded to a single colony on the original MS agar plate.

**Confirmation of positive colony**

5 The single colony corresponding to the strongly hybridising spot was inoculated into 10 ml TSB:S-YEME liquid medium, and grown for 40 h at 30°C. Mycelia were harvested and washed, and used as a template for PCR with primers "SC7E4.19" and "TN4560R-2", as previously  
10 described. A PCR product of the same size (~1.7 kb, as determined by agarose gel electrophoresis) as was observed with genomic DNA from minilibrary 3 was obtained. Sequencing of the PCR product, as described previously, confirmed that we have isolated the specific mutant strain  
15 of interest, with Tn4560 inserted in SC7E4.20, at position 18686..18690.

**Example 3****Locations of Tn4560 insertions in randomly selected clones of *S. coelicolor* M145**

5 The aim of the experiments presented in this example was to determine whether Tn4560 had preferred target sequences. For this, 41 viomycin-resistant clones were randomly picked, and the Tn4560 insertion sites were amplified using ligation-mediated PCR (LMPPCR) and 10 sequenced. The obtained sequences include the Tn4560-chromosome junctions, confirming that the correct DNA fragment was amplified.

15 The insertion sites are compiled in **Table 1**, which lists the coordinates of the insertion sites (DR coordinates) in the cosmid clone sequences which are deposited in the EMBL DNA sequence library and available also at [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/). The name of the cosmid clone is evident from the name of the gene 20 that has been disrupted or from the Description column (see example in Footnote 3 of Table 1). The 5 bp Tn4560 target sequences are tandemly repeated and flank the transposon as a result of transposition.

25 Tn4560 insertions were found all over the 8.7 Mb linear *S. coelicolor* chromosome. Gaps without Tn4560 insertion may have been missed purely by chance. Four out of the 41 insertions are between open reading frames, entirely compatible with a random distribution between coding and 30 non-coding sequences. Genes of different functions, including known named genes and novel unknown genes, have been disrupted. No common features were recognised in the target sequences. In particular, there was no specific consensus target sequence or G+C bias.

The data are compatible with, but do not prove, random transposition of Tn4560 in *S. coelicolor* M145. They confirm that Tn4560 is useful for the disruption of a great variety of *S. coelicolor* genes. Of course, the technique does not allow the recovery of transposon insertions in genes that were essential at any stage during library construction. The number of these essential genes is unknown, but not expected to exceed 10% of all genes, leaving about 7000 genes amenable to Tn4560 mutagenesis.

**LMPCR (Ligation-mediated PCR) amplification of Tn4560-chromosome junctions**

LMPCR was described first by Mueller and Wold (1989). LMPCR relies on the ligation of a synthetic, partially double-stranded, non-phosphorylated oligonucleotide adapter to genomic DNA. PCR amplification then occurs between a transposon-specific primer and an adapter-specific primer. Successful amplification requires the presence of a suitable restriction site within about 2 kb of the Tn4560 insertion. The restriction endonuclease *Eag*I (C↓GGCCG) which cleaves the high G+C *S. coelicolor* genome about 25,000 times was used preferentially for digesting the genomic DNA. If no amplification was achieved, *Eag*I was replaced by *Ngo*AIV (G↓CCGGC) or *Apa*I (GGGCC↓C), and LMPCR was repeated.

Total DNA (0.2 - 0.3 µg) isolated from a liquid culture inoculated with spores derived from a single colony was digested overnight with the chosen restriction endonuclease (e.g. *Eag*I), in a total volume of 20 µl. The restriction endonuclease was then heat inactivated according to manufacturer's instructions. 5 µl of the

digest was ligated to 25 pmol of the adapter, in a total volume of 20  $\mu$ l, at 12°C overnight. After ligation, the T4 DNA ligase was inactivated by incubation at 65°C for 10 min.

5

For each restriction endonuclease the specific, partially double-stranded adaptor shown in Table 2 was ligated to the genomic DNA fragments. The adapters were designed to ligate to the cohesive ends of the chosen restriction 10 enzyme (e.g. *EagI*), and to eliminate the restriction site upon ligation.

Partially double-stranded adapters were prepared by annealing two complementary non-phosphorylated 15 oligonucleotides. The two oligonucleotides were mixed in equimolar amounts (1 nmol  $\mu$ l<sup>-1</sup> each) in 1x PCR buffer (Perkin Elmer), in a total volume of 10  $\mu$ l. The mixture was heated to 80°C and then cooled to 4°C over about 1 h.

20 The longer, top strand of the adaptor was also used for the PCR reaction, together with a transposon-specific primer as described above in Example 2. The Tn4560-specific primers were placed a minimum of 100 bp from the end of the transposon so that sufficient Tn4560 sequence 25 could be obtained for unambiguous identification.

Amplification of the right end of Tn4560 generally gave single bands after the first PCR amplification. Multiple non-specific bands were observed after the first round of 30 PCR. A single specific band was, however, obtained after a second PCR amplification using 1  $\mu$ l of a 1:100 dilution (in distilled water) of the initial reaction mixture as template, the adapter-specific primer and a 3'-nested

transposon-specific primer. (Two successive PCR reactions using 3' nested primers are standard procedure for LMPCR.)

PCR reaction mix:

|    |                                               | Amount  | [Final]              |
|----|-----------------------------------------------|---------|----------------------|
| 5  | 10x PCR buffer                                | 10 µl   | 1x                   |
|    | 50% DMSO                                      | 10 µl   | 5%                   |
|    | 25mM dNTP mix                                 | 0.4 µl  | 100 µM               |
|    | Adapter primer (50 pmol µl <sup>-1</sup> )    | 1 µl    | 50 pmoles            |
| 10 | Transposon primer (50 pmol µl <sup>-1</sup> ) | 1 µl    | 50 pmoles            |
|    | AmpliTaq Gold® (5U l <sup>-1</sup> )          | 0.2 µl  | 1 unit               |
|    | Ligation mixture                              | 1 µl    | 20-30 ng genomic DNA |
|    | Distilled water                               | 76.4 µl | 100 µl total volume  |

15 PCR conditions: A PTC-100™ Thermal cycler with a hot lid (MJ Research) was used. AmpliTaq Gold® was activated by an initial 10 min incubation at 95°C. Amplification was by 35 cycles of 30 s 95°C, 30 s 57°C, and 90 s 72°C. A final incubation for 10 min at 72°C was added to ensure completion of the polymerisation reactions.

Sequencing of the LMPCR products was performed using a Tn4560 primer as described in Example 2.

**Table 1 Randomly selected Tn4560 insertions in *S. coelicolor* M145**

| Num<br>ber | 5 bp DR <sup>1</sup> | DR co-<br>ordina-<br>tes <sup>2</sup> | Gene <sup>3</sup> | Description <sup>4</sup>                                       |
|------------|----------------------|---------------------------------------|-------------------|----------------------------------------------------------------|
| 01         | GTGCT                | 32410                                 | J4.48             | probable integral membrane permease                            |
| 02         | TGGTC                | 03851                                 | <i>xyIA</i>       | xylose isomerase, 2G11.032c                                    |
| 03         | TGTTC                | 14551                                 | Intergenic        | between L6.12c and L6.13c                                      |
| 04         | GAGCC                | 13871                                 | I35.15c           | possible transcriptional regulator                             |
| 05         | AACAC                | 19596                                 | I35.19c           | probable rRNA methylase                                        |
| 06         | TGAAT                | 13985                                 | Intergenic        | between I35.15c and I35.16c                                    |
| 07         | GGATA                | 13547                                 | I35.15c           | possible transcriptional regulator                             |
| 08         | GTCTA                | 22628                                 | I35.23            | unknown hypothetical                                           |
| 09         | TACGC                | 01928                                 | I33.02            | possible oxidoreductase                                        |
| 10         | GAGTA                | 33406                                 | <i>cobN</i>       | cobalamin biosynthesis protein, I8.34                          |
| 11         | GAGAA                | 21143                                 | I39.22            | possible secreted penicillin binding protein                   |
| 12         | TTCAT                | 13976                                 | 10B7.16           | hypothetical, similar to <i>M. tuberculosis</i>                |
| 13         | TACCA                | 07638                                 | C8A.07            | hypothetical                                                   |
| 14         | AAGCC                | 03315                                 | C24.03c           | unknown hypothetical                                           |
| 15         | GACCT                | 23922                                 | 6D10.16           | possible sugar transporter inner membrane protein              |
| 16         | ATCGA                | 07959                                 | E34.05c           | possible membrane protein                                      |
| 17         | AGACA                | 11822                                 | E39.19            | putative transport protein                                     |
| 18         | GGAGC                | 01247                                 | E9.01c            | <i>gntR</i> family regulator                                   |
| 19         | GGTCT                | 04865                                 | E65.05c           | hypothetical, weakly similar to <i>M. tuberculosis</i>         |
| 20         | ACCTT                | 02896                                 | H5.16c            | hypothetical                                                   |
| 21         | AAGAA                | 02897                                 | H69.04c           | IS469 transposase, identical to E65.26                         |
| 22         | GACCC                | 02043                                 | 2D46.04           | unknown, similar to hypothetical from <i>M. tuberculosis</i> . |
| 22         | AGACC                | 31324                                 | E65.26            | probable 3-hydroxyacyl-CoA dehydrogenase                       |
| 23         | AATCC                | 27370                                 | Intergenic        | possible GntR-family transcriptional regulator                 |
| 24         | GACCC                | 10709                                 | K20.11c           | IS466A transposase, identical to E65.05c                       |
| 25         | AACTC                | 14408                                 | K13.13            | between D10.26 and SCD10.27                                    |
| 26         | ACCTC                | 27912                                 | K7.23             | probable monooxygenase                                         |
| 27         | GGGTC                | 16060                                 | K7.13c            | unknown                                                        |
| 28         | TACTA                | 17421                                 | 7E4.19            | possible nucleotide sugar dehydrogenase                        |
| 29         | TAGAC                | 04437                                 | 6A11.04c          | conserved hypothetical                                         |
| 30         | GACTC                | 35668                                 | 6A9.38            | sesquiterpene synthase                                         |
| 31         | GAGTC                | 14473                                 | Intergenic        | possible integral membrane transport protein                   |
| 32         | GCACT                | 24017                                 | <i>agaY</i>       | putative ATP/GTP-binding protein                               |
| 33         | TTCAT                | 19531                                 | <i>redW</i>       | between 7C7.03 and 7C7.04                                      |
| 34         | GATCA                | 05018                                 | 1B5.05c           | tagatose-bisphosphate aldolase, 9B10.19                        |
| 35         | GAGCC                | 12768                                 | 1A6.11            | acyl-CoA dehydrogenase, 2E9.20                                 |
| 36         | GAACC                | 22321                                 | 9C7.20            | unknown, slight similarity to <i>H. influenzae</i>             |
| 37         | AGGCT                | 10338                                 | 4B5.08c           | hypothetical                                                   |
| 38         | GTCAC                | 16805                                 | 4C6.16            | probable two-component regulator, <i>luxR/uhpA</i> family      |
| 39         | GGGGA                | 30604                                 | 2H12.29           | conserved hypothetical                                         |
| 40         | GGACT                | 20630                                 | 10G8.24c          | oxidoreductase                                                 |
| 41         | GTCCC                | 23334                                 | 6D11.21           | probable glutathione-dependent aldehyde dehydrogenase          |
|            |                      |                                       |                   | possible two component system response regulator               |
|            |                      |                                       |                   | possible secreted protein                                      |
|            |                      |                                       |                   | possible multidrug-efflux transporter protein                  |

**Footnotes for Table 1**

<sup>1</sup> 5 bp DR (direct repeat) of genomic sequence normally generated on Tn4560 insertion.

<sup>2</sup> Coordinates of the leftmost base of the 5 bp DR in the clone sequences which are deposited in the EMBL DNA sequence library. (Note that DNA sequences may occasionally be updated to correct rare sequencing errors.)

<sup>3</sup> *S. coelicolor* genes have multiple, synonymous names. When no functional designation (e.g. *xylA*) was available, the name of the cosmid clone and the number of the gene are given (e.g. J4.48 means gene number 48 on cosmid SCJ4 (SC indicating that the clone is from *S. coelicolor*); c after the gene number indicates that the gene is encoded by the complementary strand).

<sup>4</sup> The descriptions are taken from the annotation of the EMBL files.

**Table 2. Restriction endonucleases and corresponding adaptors for LMPCR**

| Enzyme          | Adapter                                              |
|-----------------|------------------------------------------------------|
| <i>Apa</i> I    | GTCTCTGAATTACACATCTAGGCC<br>       <br>TTAAGTGTAGAT  |
| <i>Eag</i> I    | GACTCGCGAATTCCGACAGTTGA<br>       <br>GCTGTCAACTCCGG |
| <i>Ngo</i> AIIV | GACTCGCGAATTCCGACAGTTGA<br>       <br>GCTGTCAACTGGCC |

**References**

**Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1989)** Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, NY.

**Bennett PM, Grinsted J, Richmond MH (1977)** Mol. Gen. Genet., 154, 205-211.

**Berg DE, Howe MN (1989)** (eds) "Mobile DNA" American Society for Microbiology, Washington, DC, pages 163-184.

**Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schoner BE (1992)** Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene, 116, 43-49.

**Birch AW, Cullum J (1985)** Temperature-sensitive mutants of the *Streptomyces* plasmid pIJ702. J. Gen. Microbiol., 131, 1299-1303.

**Brolle DF, Pape H, Hopwood DA, Kieser T (1993)** Analysis of the transfer region of the *Streptomyces* plasmid SCP2. Mol. Micro., 10, 157-170.

**Chung S-T (1987)** Tn4556, a 6.8-kilobase-pair transposable element of *Streptomyces fradiae*. J. Bacteriol., 169, 4436-4441.

**Chung S-T, Crose LL (1990)** Transposon Tn4556 mediated DNA insertion and site-directed mutagenesis. In 6<sup>th</sup> International Symposium on Genetics of Industrial Microorganisms, GIM 90, Proceedings. Heslot H, Davies J, Florent J, Bobichon L, Durnad

G, Penasse L (Eds). Societe Française de Microbiologie, pp. 207-218.

Flett F, Mersinias V, Smith CP (1997) *FEMS Microbiol. Lett.*, 155, 223-229.

Grindley ND, Read RR (1985) *Annual Review of Biochemistry*, 54, 863-896.

Guiney DG, Yakobson E (1983) Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proc. Natl. Acad. Sci. USA*, 80, 3595-3598.

Katz E, Thompson CJ, Hopwood DA (1983) Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. *J. Gen. Microbiol.*, 129, 2703-2714.

Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical *Streptomyces* Genetics. *The John Innes Foundation, Norwich, UK*.

Kieser T, Hopwood DA, Wright HM, Thompson CJ (1982) pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. *Mol. Gen. Genet.*, 185, 223-238.

MacNeil DJ, Gewain GM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, 111, 61-68.

Mueller PR, Wold B (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 246, 780-786.

Olson ER, Chung ST (1988) Transposon Tn4556 of *Streptomyces fradiae*: nucleotide sequence of the ends and the target sites. *J. Bacteriol.*, 170, 1955-1957.

Pansegrouw W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Stanisich VA, Thomas CM (1994) *J.Mol.Biol.*, 239, 623-663.

Pospiech A, Neumann B (1995) A versatile quick-prep of genomic DNA from gram-positive bacteria. *Trends Genet.*, 11, 217-218.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A Laboratory Manual, 2<sup>nd</sup> ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Siemieniak DR, Slightom JL, Chung, ST (1990) Nucleotide sequence of *Streptomyces fradiae* transposable element Tn4556: a class-II transposon related to Tn3. *Gene*, 86, 1-9.

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33, 103-119.

CLAIMS:

1. A method of generating a mutation in a bacterial host cell, the method comprising:
  - (a) providing a bacterial donor cell comprising a plasmid which comprises
    - a transposable element encoding functions to enable transposition of the transposable element into the host cell nucleic acid, the transposable element further comprising a first marker gene, and
    - an origin of transfer; and
  - (b) introducing the plasmid from the donor cell to the host cell by conjugation.
2. The method of claim 1, wherein the origin of transfer is an *oriT* which can be mobilised by the helper plasmids pUZ8002 and pUB307.
3. The method of claim 2, wherein the origin of transfer is an *oriT* from an IncP-group plasmid.
4. The method of claim 3, wherein the origin of transfer has the nucleic acid sequence:

CCGGGCAGGA TAGGTGAAGT AGGCCACCC GCGAGCGGGT GTTCCTTCTT  
CACTGTCCCT TATT CGCACC TGGCGGTGCT CAACGGGAAT CCTGCTCTGC  
GAGGCTGGC,

or a variant thereof having at least 50% nucleic acid sequence identity thereto and having origin of transfer function.
5. The method of any preceding claim, wherein the plasmid comprises an origin of replication which is functional in the donor cell.

6. The method of any preceding claim, wherein the host cell lacks plasmids that accept transposons.
7. The method of any preceding claim, wherein the host cell is an actinomycete cell.
8. The method of claim 7, wherein the host cell is a cell of the genus *Streptomyces*.
9. The method of claim 8, wherein the host cell is of the species *S. coelicolor*.
10. The method of claim 9, wherein the host cell is of the strain *S. coelicolor A3(2)*.
11. The method of claim 10, wherein the host cell lacks the SCP1 and SCP2 plasmids.
12. The method of any one of claims 7 to 11, wherein the conjugation of step (b) is into pre-germinated spores.
13. The method of any preceding claim, wherein the donor cell is of a different cell type from the host cell.
14. The method of any preceding claim, wherein the donor cell is an *E.coli* cell.
15. The method of any preceding claim, wherein the donor cell is methylation-deficient.
16. The method of claim 15, wherein the donor cell is of *E.coli* strain ET12567.

17. The method of any preceding claim, wherein the transposable element undergoes replicative transposition.
18. The method of any preceding claim, wherein the transposable element also comprises a resolvase gene and internal resolution site.
19. The method of any preceding claim, wherein the transposable element confers transposition immunity following transposition into a replicon.
20. The method of any preceding claim, wherein the transposable element does not demonstrate high target site specificity.
21. The method of claim 20, wherein the transposable element has inverted repeat sequences which do not display a significant preference for particular insertion sites ("insertion hotspots") in the host cell.
22. The method of claim 20 or claim 21, wherein the transposable element does not demonstrate detectable target site specificity in the host cell.
23. The method of any preceding claim, wherein the transposable element comprises a transposase gene having at least 65% identity at the amino acid level with the transposase of Tn3 (GenBank V00613) and/or Tn4556 (GenBank M29297).
24. The method of claim 23, wherein the level of identity is at least 80%.

25. The method of any preceding claim, wherein the transposable element comprises a resolvase gene having at least 45% identity at the amino acid level with the resolvase of Tn3 (GenBank V00613) and/or Tn4556 (GenBank M29297).
26. The method of claim 25, wherein the level of identity is at least 65%.
27. The method of claim 26, wherein the level of identity is at least 85%.
28. The method of any preceding claim, wherein the transposable element comprises inverted repeat sequences having at least 70% identity at the nucleic acid level with the inverted repeat sequences of Tn3 (GenBank V00613) and/or Tn4556 (GenBank M29297).
29. The method of claim 28, wherein the level of identity is at least 85%.
30. The method of any preceding claim, wherein the plasmid comprises, outside the transposable element, a second marker gene which differs from the first marker gene.
31. The method of any preceding claim, wherein the or a marker gene is an antibiotic or heavy metal resistance gene.
32. The method of any preceding claim, wherein the plasmid is non-replicable in the host cell.
33. The method of claim 32, wherein plasmid lacks a functional origin of replication for the host cell.

34. The method of any preceding claim, wherein the plasmid is pUC1169::pOJ260 or a derivative thereof.

35. The method of claim 34, wherein the plasmid is pKay1.

36. The method of any preceding claim, wherein step (b) includes the provision *in trans* of transfer function.

37. The method of claim 36, wherein the transfer function is provided by an *E.coli* donor strain carrying the self-transmissible pUB307 or the non-transmissible pUZ8002.

38. The method of claim 36 or claim 37, wherein step (b) comprises transforming the plasmid into a donor strain carrying a non-transmissible transfer plasmid, followed by incubation under suitable conditions with the host cell.

39. The method of any preceding claim, wherein the conjugation step (b) comprises an incubation step or steps to allow conjugation and subsequently transposition to occur.

40. The method of claim 39, comprising incubation between about 30°C and 37°C for transposition to occur.

41. The method of any preceding claim which is carried out simultaneously on several host cells.

42. The method of claim 41, wherein following conjugation and transposition the host cells are pooled to produce a library of independently mutated host cells.

43. The method of any preceding claim, wherein the conjugation step (b) is followed by a selective incubation step (c) under conditions which select for the host cell in

preference to the donor cell and for the presence in the host cell of the first marker gene.

44. The method of claim 43, wherein the selective incubation step is followed by a step (d) of identifying host cells which possess the first marker gene, but which lack the second marker gene, wherein possession of the first marker gene and lack of the second marker gene is indicative of a transposition event having occurred.

45. The method of any preceding claim, further comprising the step of storing the resultant mutant host cells.

46. A host cell producible or as produced by the process of any preceding claim.

47. A method for producing a library of independently mutated host cells, the method comprising:

(a) providing a plurality of bacterial donor cells each comprising a plasmid which comprises

a transposable element encoding functions to enable transposition of the transposable element into the host cell nucleic acid, the transposable element further comprising a first marker gene, and

an origin of transfer; and

(b) introducing plasmids from respective donor cells into respective host cells by conjugation;

(c) optionally carrying out a step of selective incubation under conditions which select for host cells in preference to donor cells and for the presence in the host cell of the first marker gene;

(d) optionally identifying host cells which possess the first marker gene, but which lack the second marker gene, wherein possession of the first marker gene and lack of the

second marker gene is indicative of a transposition event having occurred; and

pooling the host cells to form a library of host cells..

48. The method of claim 47, wherein steps (a) to (d) are as defined in any one of claims 1 to 45.

49. The method of claim 48, wherein the library comprises at least 50 independent transposon mutant host cells.

50. The method of claim 49, wherein the library comprises at least 1000 independent transposon mutant host cells.

51. The method of claim 50, wherein the library comprises at least 50000 independent transposon mutant host cells.

52. A library of independently mutated host cells as produced or as producible according to any one of claims 47 to 51..

53. A plasmid for generating a genetic disruption in a bacterial host cell, the plasmid comprising:

a transposable element encoding functions to enable transposition of the transposable element into the host cell nucleic acid, the transposable element further comprising a first marker gene, and

an origin of transfer.

54. The plasmid of claim 53, which is as defined in any one of claims 2 to 5 and 17 to 35.

55. A donor cell comprising a plasmid according to claim 53 or claim 54.

56. The donor cell of claim 55, which is as defined in any one of claims 14 to 16.

57. A kit comprising the plasmid of claim 53 or claim 54, a bacterial donor cell and a transfer plasmid capable of providing transfer function *in trans*.

58. The kit of claim 57, wherein donor cell contains the plasmid of claim 53 or claim 54 and optionally transfer plasmid, or these components may be provided separately.

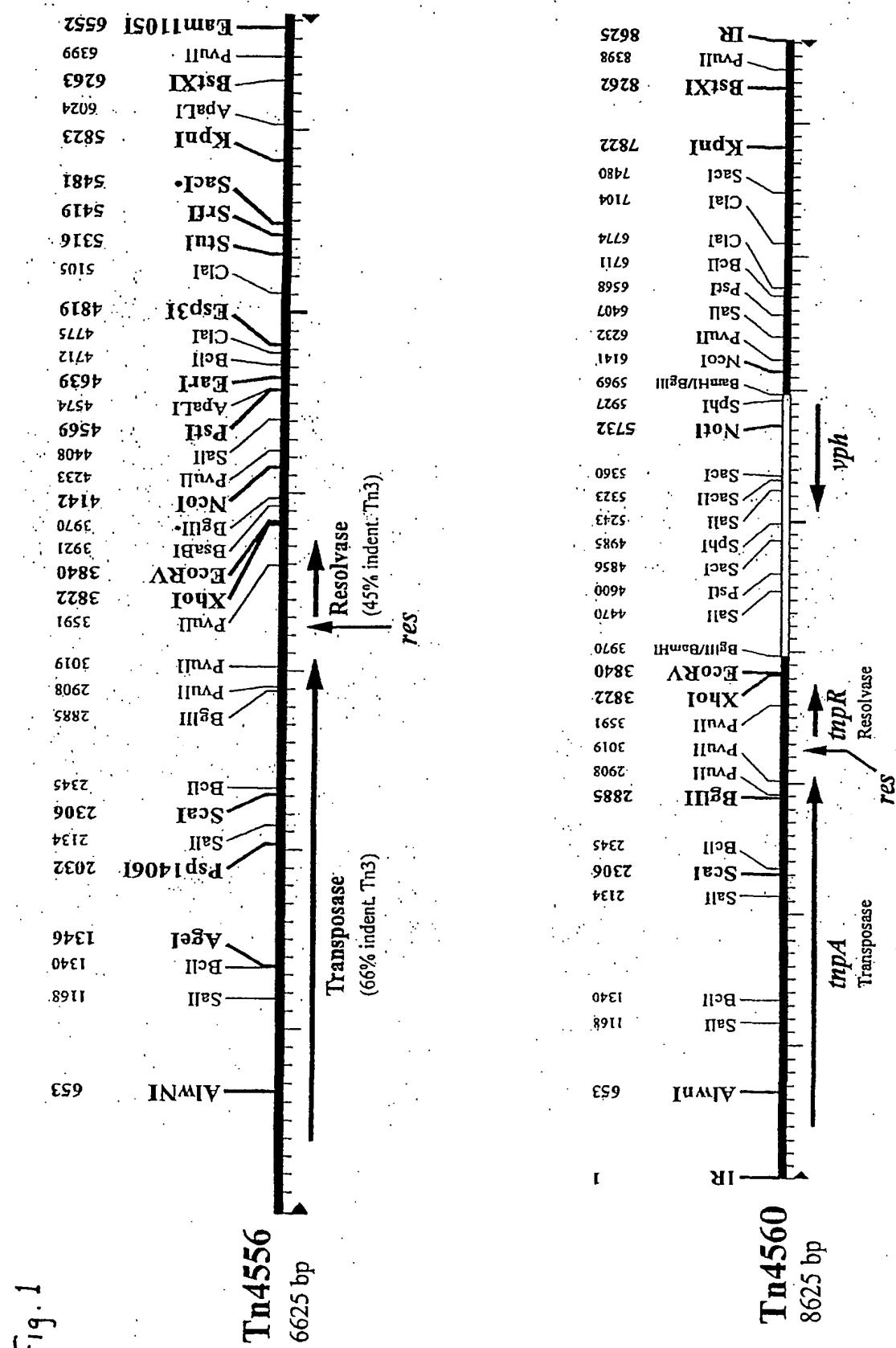


Fig. 1

Fig. 2

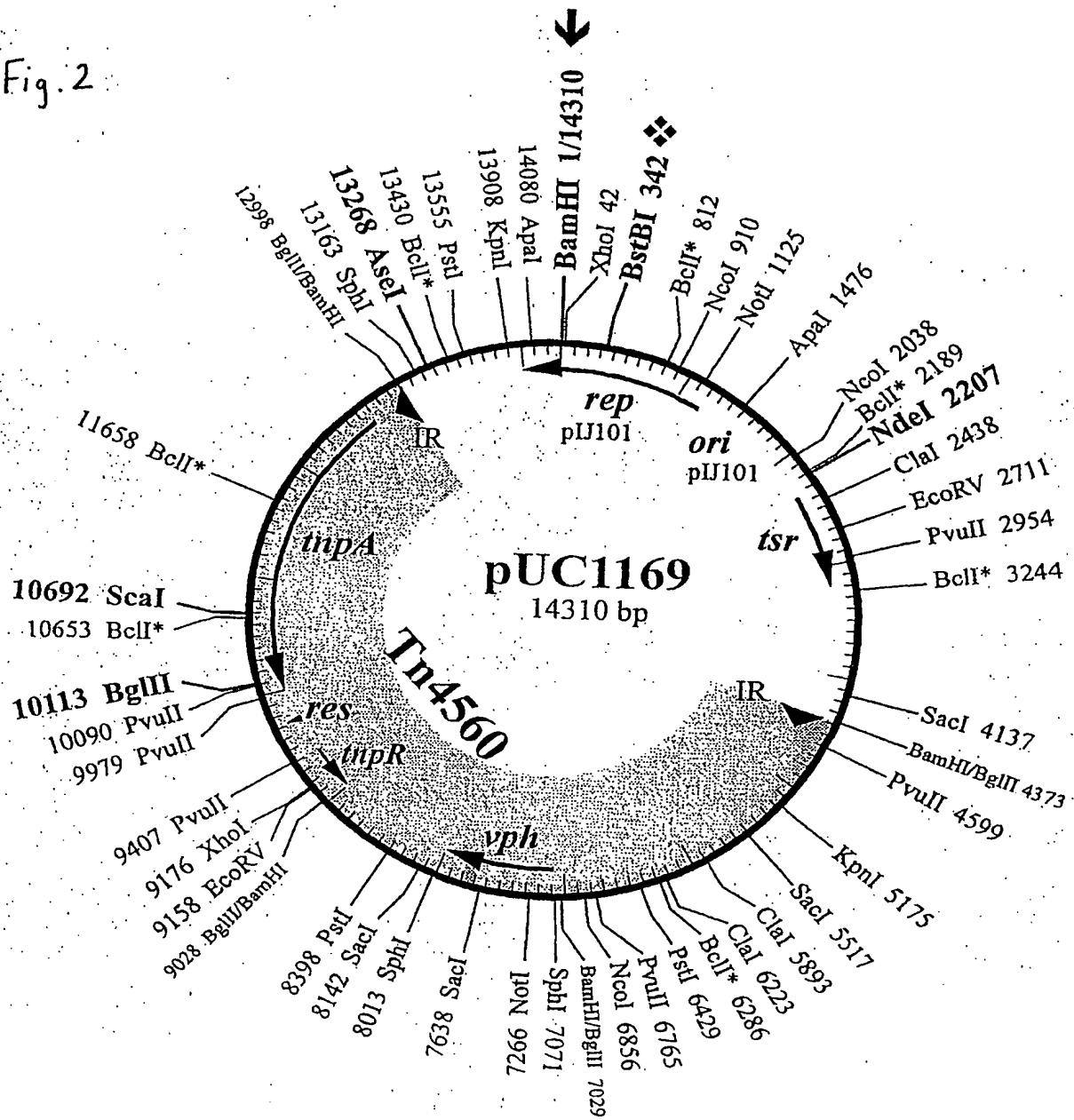


Fig. 3

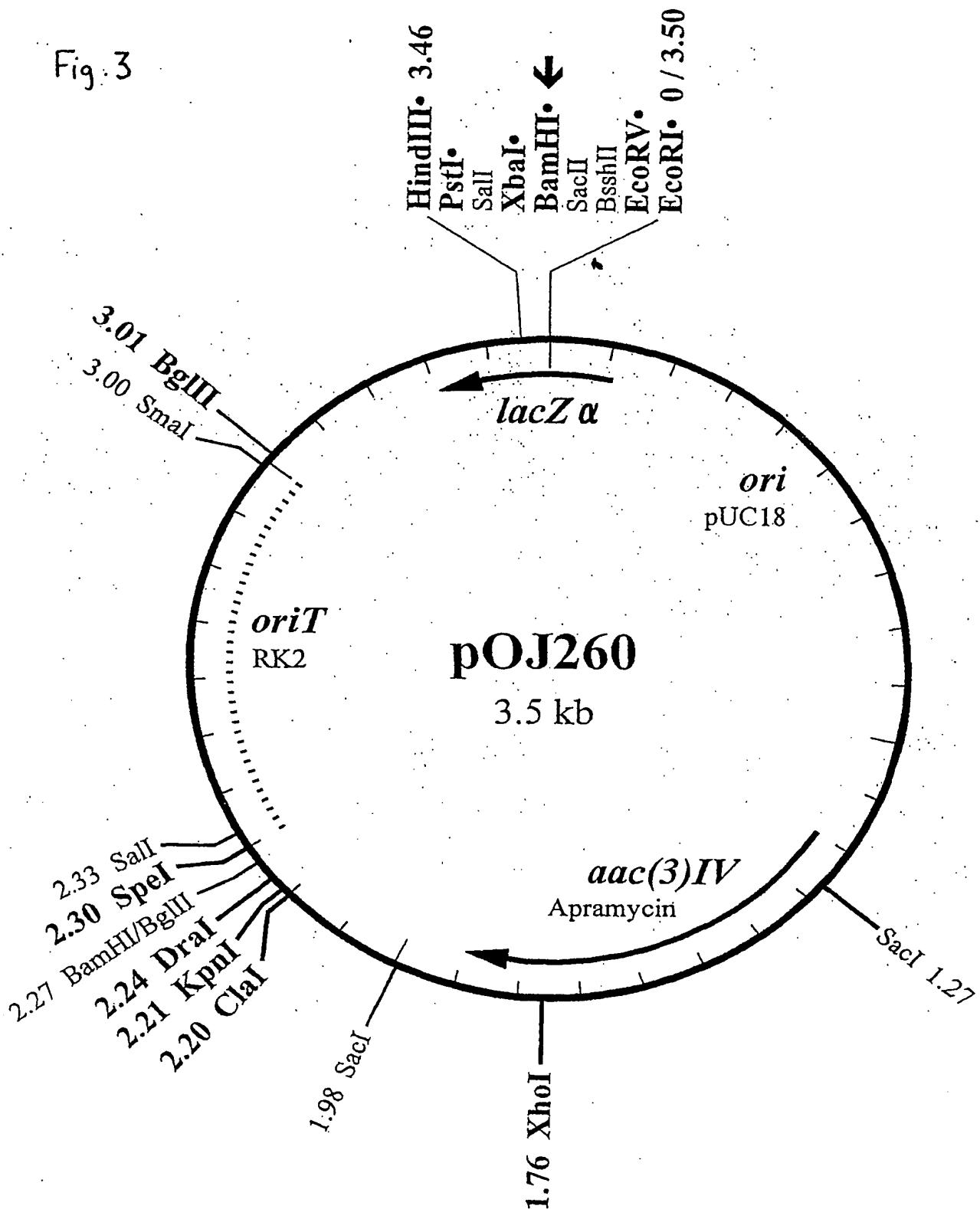


Fig. 4a

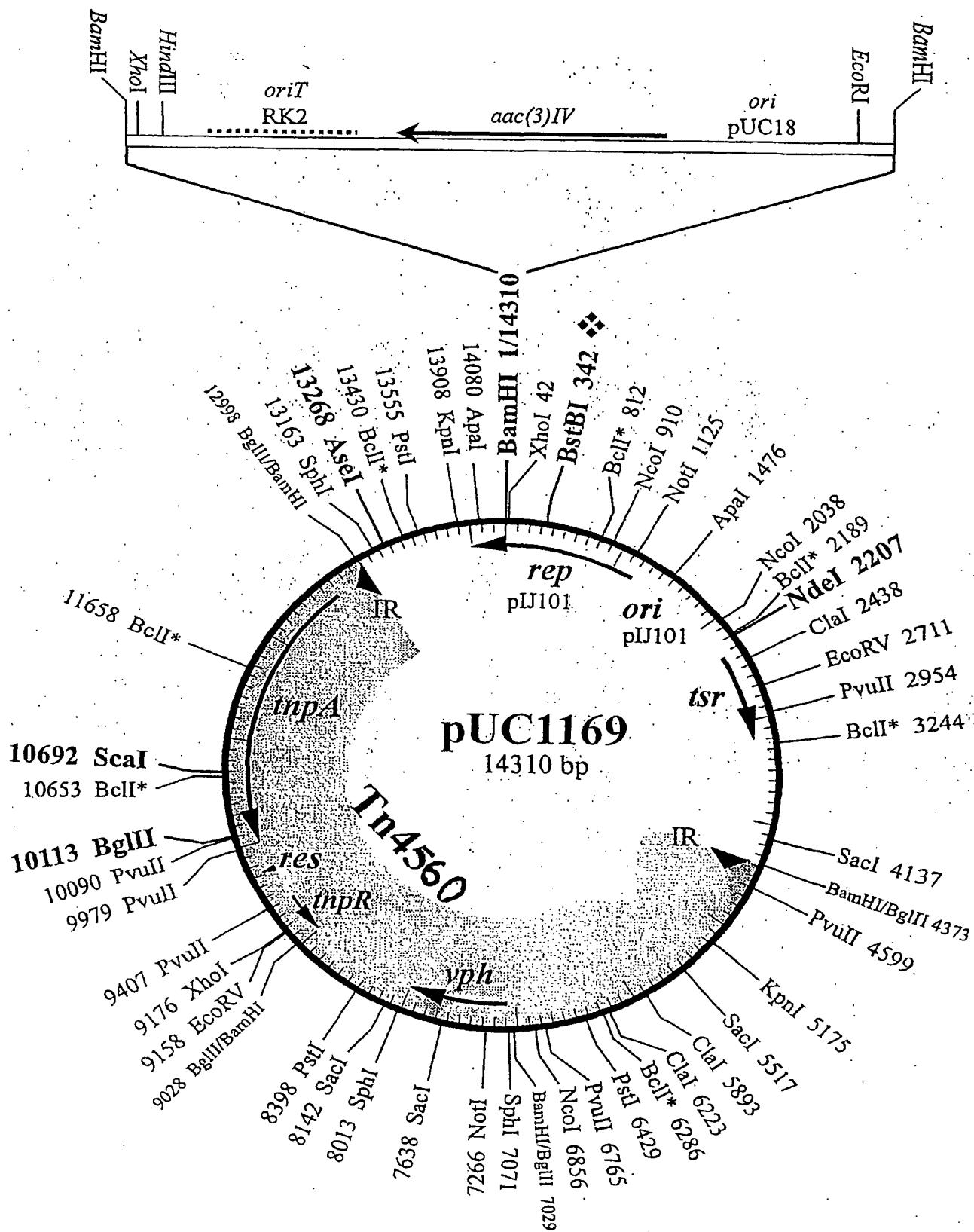


Fig. 4b

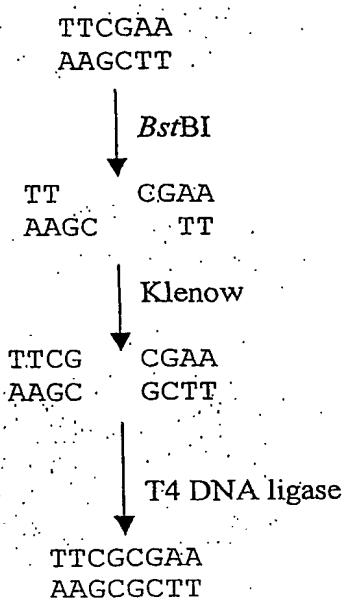


Fig. 5

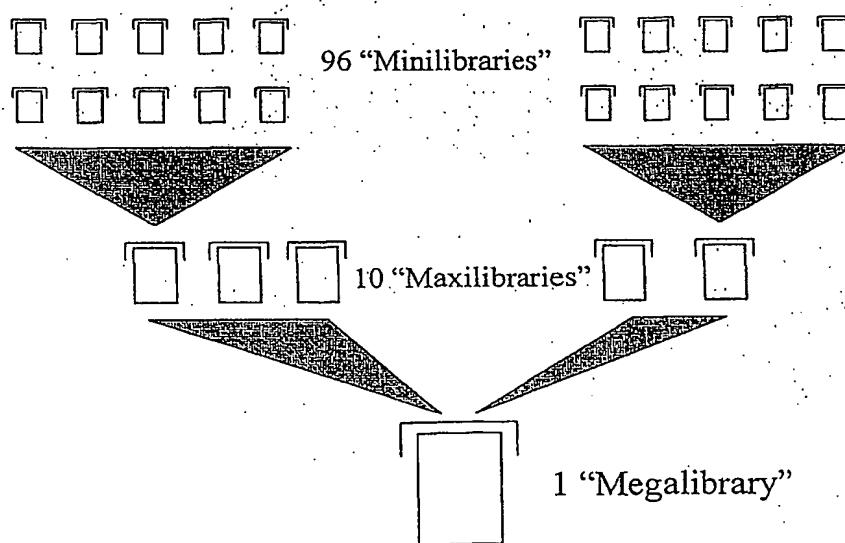
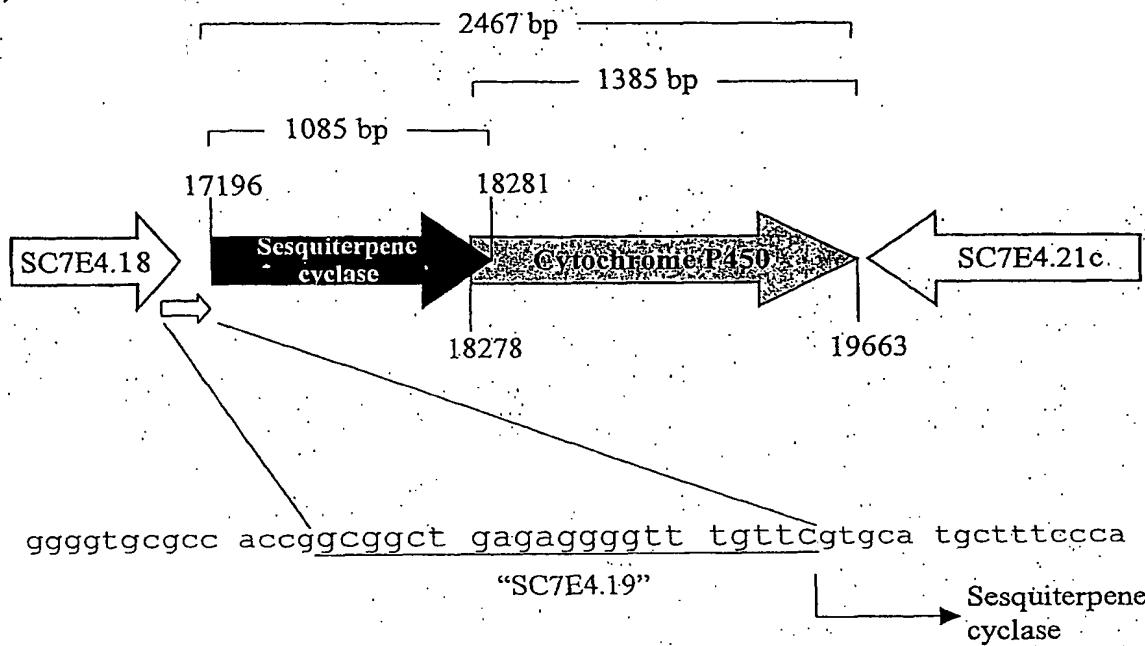
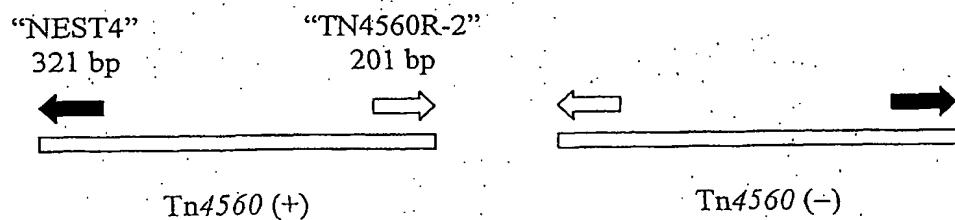


Fig. 6

(a)



(b)



## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 02/02884

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12N15/10 C12N15/76 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, MEDLINE, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages             | Relevant to claim No.                                                              |
|----------|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| X        | CA 1 264 685 A (ALLELIX INC)<br>23 January 1990 (1990-01-23)<br><br>the whole document<br>---- | 1, 2, 5,<br>13, 14,<br>17,<br>20-22,<br>30-33,<br>39-41,<br>43-46,<br>53-56<br>-/- |

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the International filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the International filing date but later than the priority date claimed

- \*T\* later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

|                                                                                                                                                |                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| Date of the actual completion of the International search                                                                                      | Date of mailing of the International search report |
| 21 October 2002                                                                                                                                | 18/11/2002                                         |
| Name and mailing address of the ISA                                                                                                            | Authorized officer                                 |
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl<br>Fax: (+31-70) 340-3016 | Kania, T                                           |

## INTERNATIONAL SEARCH REPORT

I      International Application No  
PCT/GB 02/02884

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                          |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Category                                             | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                  | Relevant to claim No.                                                                    |
| X                                                    | EP 0 445 385 A (DEGUSSA)<br>11 September 1991 (1991-09-11)<br><br>the whole document                                                                                                                                                                                                                                                                                                | 1, 2, 5,<br>13, 14,<br>17,<br>20-22,<br>32, 33,<br>36,<br>39-41,<br>43, 45,<br>46, 53-58 |
| X                                                    | BIERMAN M ET AL: "Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp"<br>GENE, ELSEVIER BIOMEDICAL PRESS.<br>AMSTERDAM, NL,<br>vol. 116, no. 1, 1 July 1992 (1992-07-01),<br>pages 43-49, XP002116076<br>ISSN: 0378-1119<br>cited in the application<br>the whole document                                                          | 1-5, 7, 8,<br>12-14,<br>17,<br>31-33,<br>36,<br>38-41,<br>46, 53-58                      |
| X                                                    | GEHRING ET AL: "Genomewide insertional mutagenesis in Streptomyces coelicolor reveals additional genes involved in morphological differentiation"<br>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US,<br>vol. 97, no. 17,<br>15 August 2000 (2000-08-15), pages 9642-9647, XP002210219<br>ISSN: 0027-8424<br>the whole document | 1-17,<br>20-22,<br>30-33,<br>36-50,<br>52-58                                             |
| Y                                                    | T. KIESER: "Practical Streptomyces Genetics"<br>2000 , THE JOHN INNES FOUNDATION , NORWICH<br>XP002217484<br>cited in the application<br>page 109 -page 123                                                                                                                                                                                                                         | 18, 19,<br>23-29,<br>34, 35, 51                                                          |
| A                                                    | CHUNG S-T: "TN-4556 A 6 8-KILOBASE-PAIR TRANSPOSABLE ELEMENT OF STREPTOMYCES-FRADIAE"<br>JOURNAL OF BACTERIOLOGY,<br>vol. 169, no. 10, 1987, pages 4436-4441,<br>XP001107268<br>ISSN: 0021-9193<br>cited in the application<br>table 1                                                                                                                                              | 34                                                                                       |
|                                                      |                                                                                                                                                                                                                                                                                                                                                                                     | -/-                                                                                      |

## INTERNATIONAL SEARCH REPORT

II Final Application No  
PCT/GB 02/02884

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                       | Relevant to claim No. |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A        | FLETT FIONA ET AL: "High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes"<br>FEMS MICROBIOLOGY LETTERS, AMSTERDAM, NL,<br>vol. 155, no. 2,<br>15 October 1997 (1997-10-15), pages<br>223-229, XP002210221<br>ISSN: 0378-1097<br>cited in the application<br>----                 |                       |
| A        | SIEMIENIAK D R ET AL: "NUCLEOTIDE SEQUENCE OF STREPTOMYCES-FRADIAE TRANSPOSABLE ELEMENT TN4556 A CLASS-II TRANSPOSON RELATED TO TN3"<br>GENE (AMSTERDAM),<br>vol. 86, no. 1, 1990, pages 1-10,<br>XP001106585<br>ISSN: 0378-1119<br>cited in the application<br>----                                                                                     |                       |
| A        | HERRERO M ET AL: "TRANSPOSON VECTORS CONTAINING NON-ANTIBIOTIC RESISTANCE SELECTION MARKERS FOR CLONING AND STABLE CHROMOSOMAL INSERTION OF FOREIGN GENES IN GRAM-NETATIVE BACTERIA"<br>JOURNAL OF BACTERIOLOGY, WASHINGTON, DC,<br>US,<br>vol. 172, no. 11,<br>1 November 1990 (1990-11-01), pages<br>6557-6567, XP000572232<br>ISSN: 0021-9193<br>---- |                       |
| A        | BALTZ RICHARD H ET AL: "Applications of transposition mutagenesis in antibiotic producing streptomycetes."<br>ANTONIE VAN LEEUWENHOEK,<br>vol. 71, no. 1-2, 1997, pages 179-187,<br>XP001107054<br>ISSN: 0003-6072<br>-----                                                                                                                              |                       |

## INTERNATIONAL SEARCH REPORT

|                 |                      |
|-----------------|----------------------|
| Int             | Int'l Application No |
| PCT/GB 02/02884 |                      |

| Patent document cited in search report |   | Publication date |    | Patent family member(s) |  | Publication date |
|----------------------------------------|---|------------------|----|-------------------------|--|------------------|
| CA 1264685                             | A | 23-01-1990       | CA | 1264685 A1              |  | 23-01-1990       |
| EP 0445385                             | A | 11-09-1991       | DE | 4006637 A1              |  | 05-09-1991       |
|                                        |   |                  | AT | 93274 T                 |  | 15-09-1993       |
|                                        |   |                  | CA | 2037431 A1              |  | 04-09-1991       |
|                                        |   |                  | DE | 59002394 D1             |  | 23-09-1993       |
|                                        |   |                  | DK | 445385 T3               |  | 11-10-1993       |
|                                        |   |                  | EP | 0445385 A2              |  | 11-09-1991       |
|                                        |   |                  | ES | 2058738 T3              |  | 01-11-1994       |
|                                        |   |                  | JP | 3148262 B2              |  | 19-03-2001       |
|                                        |   |                  | JP | 6153925 A               |  | 03-06-1994       |
|                                        |   |                  | US | 5516669 A               |  | 14-05-1996       |